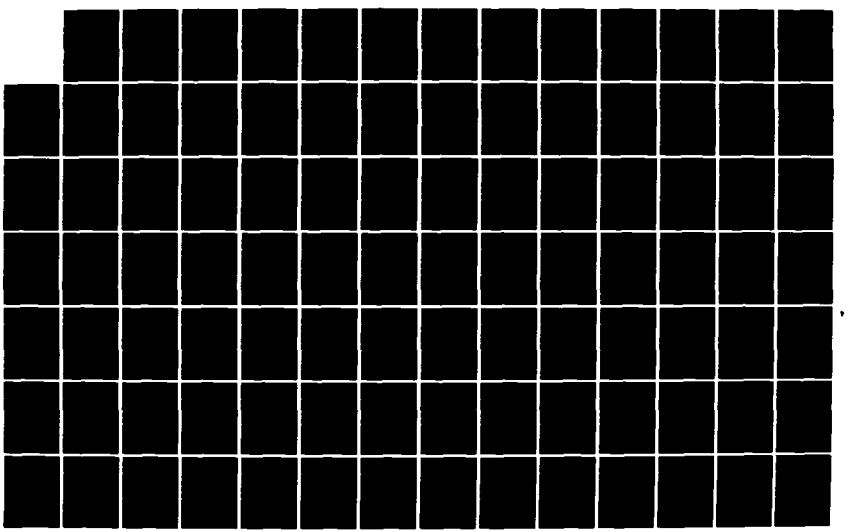
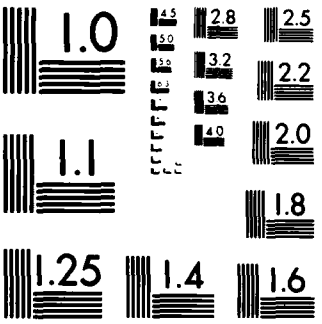
AD-A130 393 THE DETECTION OF SIGNALS IN IMPULSIVE NOISE(U) 1/2
PRINCETON UNIV NJ INFORMATION SCIENCES AND SYSTEMS LAB
E J MODUGNO ET AL. JUN 83 TR-13 N00014-81-K-0146
UNCLASSIFIED F/G 9/4 NL





REPORT NUMBER 13

THE DETECTION OF SIGNALS IN IMPULSIVE NOISE

E.J. Modugno III and J.B. Thomas

INFORMATION SCIENCES AND SYSTEMS LABORATORY

Department of Electrical Engineering and Computer Science
Princeton University
Princeton, New Jersey 08544

JUNE 1983

Prepared for

OFFICE OF NAVAL RESEARCH (Code 411SP)
Statistics and Probability Branch
Arlington, Virginia 22217
under Contract N00014-81-K-0146
SRO(103) Program in Non-Gaussian Signal Processing

S.C. Schwartz, Principal Investigator

Approved for public release; distribution unlimited

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 13	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) THE DETECTION OF SIGNAL IN IMPULSIVE NOISE		5. TYPE OF REPORT & PERIOD COVERED Technical Report Sept. 1981-February 1982 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) E.J. Modugno III and J.B. Thomas		8. CONTRACT OR GRANT NUMBER(s) N00014-81-K-0146
9. PERFORMING ORGANIZATION NAME AND ADDRESS Information Sciences and Systems Lab. Dept. of Electrical Eng. and Computer Sci. Princeton Univ., Princeton, NJ 08544		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR SRO 103
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research (Code 411 SP) Department of the Navy Arlington, Virginia 22217		12. REPORT DATE June 1983 13. NUMBER OF PAGES 186
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Also submitted as the Ph.D. Thesis of E.J. Modugno III to the EECS Department, Princeton University, Princeton, NJ, 08544, February 1982.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Detection
Non-Gaussian Detection
Impulsive Noise

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This dissertation addresses the problem of detecting known, discrete-time signals in additive non-Gaussian noise. The case of statistically independent samples is emphasized. After a brief introduction to the detection problem, the characteristics and sources of impulsive noise are discussed. Several models for impulsive noise are then presented. The complexity of these models and the need for simple density functions to approximate

(con't.)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(Abstract con't.)

the first order characteristics of impulsive noise justify consideration of three systems of densities. These three systems are: a generalized Gaussian noise, the Johnson S_u System, and a mixture model. These are used throughout this dissertation to provide examples.

In many detection problems it may only be possible to define a class of probability densities which contains the actual noise density. In such cases minimax detectors may be used to guarantee a lower bound on detector performance for the entire class. The minimax detector is the optimum detector for the worst case density. It is shown that the worst case density, in terms of minimizing the asymptotic probability of detecting a signal, is that density which minimizes Fisher's Information over the entire class. Several classes of densities are considered and conditions are established for the minimax detectors. It is also shown that impulsive noise with a large value for Fisher's Information provides a better environment for detection than does a Gaussian noise of equal power. The increased structure of the impulsive noise may be exploited to increase detector performance over that possible for Gaussian noise. Nonlinear processing, however, is required.

An adaptive detector which uses the Johnson System as a noise model is then presented. After a study of different measures of tail behavior, a scheme to adapt a detector utilizing quantiles to measure the tail behavior of the first order density function of the noise is developed. Simulation results are presented which indicate that this detector will perform well over a large variety of background noises. These simulations also support the statement that non-Gaussian noise, with large value for Fisher's Information, provides a better environment for detection than does Gaussian noise.

Suggestions for continuing this research are also included.

THE DETECTION OF SIGNAL IN IMPULSIVE NOISE

E.J. Modugno III and J.B. Thomas
Department of Electrical Engineering
and Computer Science
Princeton University
Princeton, N.J. 08544

ABSTRACT

This dissertation addresses the problem of detecting known, discrete-time signals in additive non-Gaussian noise. The case of statistically independent samples is emphasized. After a brief introduction to the detection problem, the characteristics and sources of impulsive noise are discussed. Several models for impulsive noise are then presented. The complexity of these models and the need for simple density functions to approximate the first order characteristics of impulsive noise justify consideration of three systems of densities. These three systems are: a generalized Gaussian noise, the Johnson S_u System, and a mixture model. These are used throughout this dissertation to provide examples.

In many detection problems it may only be possible to define a class of probability densities which contains the actual noise density. In such cases minimax detectors may be used to guarantee a lower bound on detector performance for the entire class. The minimax detector is the optimum detector for the worst case density. It is shown that the worst case density, in terms of minimizing the asymptotic probability of detecting a signal, is that density which minimizes Fisher's Information over the entire class. Several classes of densities are considered and conditions are established for the minimax detectors. It is also shown that impulsive noise with a large value for Fisher's Information provides a better environment for detection than does a Gaussian noise of equal power. The

increased structure of the impulsive noise may be exploited to increase detector performance over that possible for Gaussian noise. Non-linear processing, however, is required.

An adaptive detector which uses the Johnson System as a noise model is then presented. After a study of different measures of tail behavior, a scheme to adapt a detector utilizing quantiles to measure the tail behavior of the first order density function of the noise is developed. Simulation results are presented which indicate that this detector will perform well over a large variety of background noises. These simulations also support the statement that non-Gaussian noise, with large values for Fisher's Information, provides a better environment for detection than does Gaussian noise.

Suggestions for continuing this research are also included.

TABLE OF CONTENTS

Chapter I

INTRODUCTION

This dissertation addresses the problem of detecting known, discrete-time signals in additive non-Gaussian noise. This problem may arise in communications, radar, sonar, and other situations where the dominant noise may be impulsive in nature. For example, low frequency radio channels are plagued by highly impulsive lightning strokes [1]. Higher frequencies must contend with unintentionally generated man-made noise [2]. Among the sources of impulsive noise on the sonar channel [3] are ice cracking and snapping shrimp. A more detailed presentation of the sources and characteristics of impulsive noise is given in Chapter II.

Throughout this dissertation the noise samples will be considered independent. Although this assumption is not in general valid [4], it allows n^{th} order noise density functions to be expressed as products of first order densities thus greatly simplifying the detection problem. For many of the results of this dissertation, the independence assumption can be relaxed using the weaker condition of strong mixing (see, for example, Chapter III - Section 1).

The detection problem, with which this dissertation is concerned, is perhaps best expressed as a hypothesis testing problem:

$$\begin{aligned} H:&\quad x_i = n_i \\ K:&\quad x_i = n_i + \theta s_i \end{aligned} \qquad \theta > 0,\ i=1,\ldots,N \tag{1.1}$$

H is the hypothesis that the observations $\{x_i\}_1^N$ consist of noise only. The alternative K is that in addition to the noise, a known signal $\{\theta s_i\}_1^N$ is present. The $\{n_i\}_1^N$ are the realizations of the random variables $\{N_i\}_1^N$ which may be formed from sampling an underlying continuous-time random noise process. Throughout this dissertation the $\{N_i\}_1^N$ are assumed to be independent random variables with densities $\{f_i(\cdot)\}_1^N$ where $f_i \in F$ some class of symmetric densities. The s_i are assumed to have their absolute values upper and lower bounded by positive constants M_1 and m_1 so that $0 < m_1 < |s_i| < M_1 < \infty$.

The optimum detector, in terms of maximizing the probability of detection while keeping the probability of false alarm below a certain level, is given by the Neyman-Pearson lemma [5]. This detector compares the likelihood ratio to some threshold which is chosen to achieve the desired probability of false alarm.

$$L(\underline{x}) = \frac{f_K(\underline{x})}{f_H(\underline{x})} \quad \begin{cases} > T_1 & \text{decide K} \\ < T_1 & \text{decide H} \end{cases} \tag{1.2}$$

The $f_K(\underline{x})$ and $f_H(\underline{x})$ are the probability density functions for the observation vector $\underline{X}$ under K and H respectively. Since the noise samples are assumed independent and the signal is known, the $f_K(\underline{x})$ and $f_H(\underline{x})$ can be written as products of the univariate noise densities:

$$\begin{aligned} f_K(\underline{x}) &= \prod_{i=1}^{N} f_i(x_i - \theta s_i) \\ f_H(\underline{x}) &= \prod_{i=1}^{N} f_i(x_i) \end{aligned} \tag{1.3}$$

The optimum detector can then be written as

$$L(\underline{x}) = \frac{\prod_{i=1}^{N} f_i(x_i - \theta s_i)}{\prod_{i=1}^{N} f_i(x_i)} \begin{cases} > T_1 & \text{decide K} \\ < T_1 & \text{decide H} \end{cases} \tag{1.4}$$

In order to simplify the detector, it is common practice to take the natural logarithm of Eq. (1.4). This yields the Neyman-Pearson optimum detector:

$$T_{opt}(\underline{x}) = \sum_{i=1}^{N} \log \frac{f_i(x_i - \theta s_i)}{f_i(x_i)} \begin{cases} > T_o & \text{decide K} \\ < T_o & \text{decide H} \end{cases} \tag{1.5}$$

Figure 1.1 displays a block diagram for this detector. The structure is that of a time dependent nonlinearity

$$g_i(x_i) = g_{opt_i}(x_i) = \log \frac{f_i(x_i - \theta s_i)}{f_i(x_i)} \tag{1.6}$$

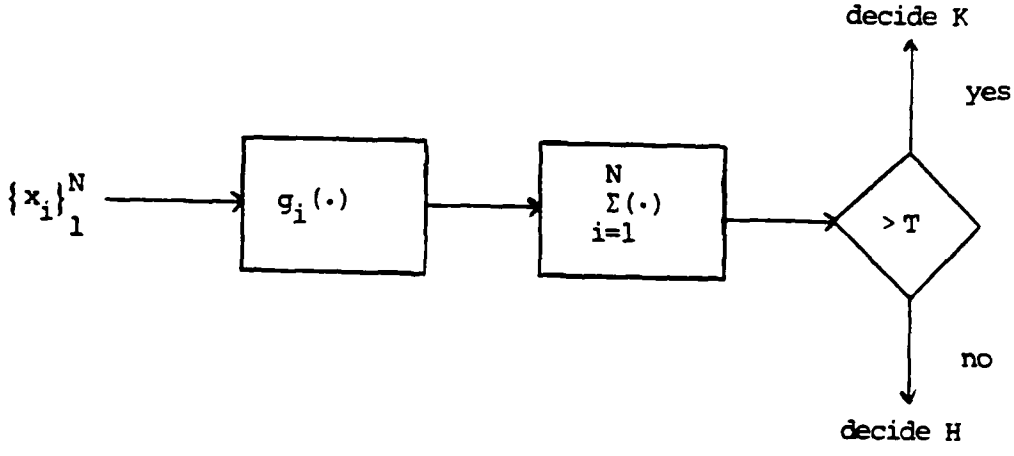


Figure 1.1: Detector Structure

followed by a summer and a threshold comparator. Unfortunately, this nonlinearity is often difficult to implement.

In small signal cases a simpler detector, the locally optimum detector, is often used. Unlike the Neyman-Pearson optimum detector, which maximizes the probability of detection (from here on referred to as the power of the detector) for a fixed probability of false alarm (referred to as the level), the locally optimum detector maximizes the slope of the power function (power versus signal strength) at the origin. Since the slope of the power function corresponds to the increase in power as signal strength is increased, the locally optimum detector's performance should be nearly as good as that of the optimum detector for small signals.

The locally optimum likelihood ratio [6] is

$$L(\underline{x}) = \left. \frac{\frac{\partial}{\partial \theta} f_K(\underline{x})}{f_H(\underline{x})} \right|_{\theta \to 0} \tag{1.7}$$

Since $f_K(\underline{x})$ approaches $f_H(\underline{x})$ as $\theta \to 0$, $L(\underline{x})$ can be written

$$L(\underline{x}) = \left. \frac{\partial}{\partial \theta} \log f_K(\underline{x}) \right|_{\theta \to 0} \tag{1.8}$$

Substituting in (1.3), taking the derivative, and allowing θ to approach zero, yields the standard form for the locally optimum detector.

$$T_{loc}(\underline{x}) = \sum_{i=1}^{N} \frac{f_i'(x_i)}{f_i(x_i)} \; (-s_i) \; \begin{cases} > T_o & \text{decide K} \\ < T_o & \text{decide H} \end{cases} \tag{1.9}$$

This detector has the same structure displayed in Figure 1.1 with

$$g_i(x_i) = g_{loc_i}(x_i) = \frac{f_i'(x_i)}{f_i(x_i)} (-s_i) \tag{1.10}$$

The time dependent nonlinearities $g_{opt_i}(\cdot)$ and $g_{loc_i}(\cdot)$ are related [7] by the equation

$$g_{opt_i}(x_i) = \frac{1}{s_i} \int_{x_i - \theta s_i}^{x_i} g_{loc_i}(t)dt \tag{1.11}$$

and the fact that $g_{loc_i}(\cdot)$ is the linear term in the Taylor expansion of $g_{opt_i}(\cdot)$:

$$g_{opt_i}(x_i) = \theta g_{loc_i}(x_i) + \{\text{higher order terms}\} \tag{1.12}$$

For small values of θ, $g_{opt_i}(\cdot)$ is often approximated by $\theta g_{loc_i}(\cdot)$.

In addition to the power function, another useful measure of detector performance is provided by the concepts of efficacy and asymptotic relative efficiency [6]. Asymptotic relative efficiency (abbreviated ARE) is used to compare two detectors based on the asymptotic ratio of the number of samples required by each to achieve the same level and power. That is, if $n_1(\alpha, \beta, \theta)$ is the number of samples required by Detector 1 to achieve level α and power β with signal strength θ and $n_2(\alpha, \beta, \theta)$ is the corresponding number for Detector 2, then

$$ARE_{1,2} = \lim_{\theta \to 0} \frac{n_2(\alpha, \beta, \theta)}{n_1(\alpha, \beta, \theta)} \tag{1.13}$$

Thus, if Detector 2 requires (in the limit) twice as many samples as Detector 1 ($n_2 = 2n_1$), then $ARE_{1,2} = 2$ and one could say Detector 1 is "twice as efficient" as Detector 2. Often the linear detector is used as a reference detector for comparing nonlinear detectors using ARE. ARE is by definition a small signal (and consequently a large sample size) concept. It can be shown [6] that the nonlinear detector which has the maximum value of ARE (using some reference such as the linear detector) is the locally optimum detector.

Although Eq. (1.13) provides the definition for ARE, the concept of efficacy provides a much simpler way to calculate it. If D_1 is a threshold detector for the problem expressed by Eq. (1.1) using the test statistic $T(\underline{x})$,

$$T(\underline{x}) \quad \begin{cases} > T & \text{decide K} \\ < T & \text{decide H} \end{cases} \tag{1.14}$$

then the efficacy [8] of D_1 is given by

$$E_{D_1} = \lim_{N \to \infty} \frac{\left[\frac{\partial}{\partial \theta} E_K[T(\underline{x})] \Big|_{\theta \to 0} \right]^2}{N Var_H[T(\underline{x})]} \tag{1.15}$$

where $E_K[T(\underline{x})]$ and $Var_H[T(\underline{x})]$ are the expectation under K and the variance under H of the test statistic. The ARE can then be calculated [8] from

$$ARE_{D_1,D_2} = \frac{E_{D_1}}{E_{D_2}} \tag{1.16}$$

Another definition which will be used frequently is Fisher's measure of information [5] for the location parameter in a density. For the problem stated in equation (1.1), Fisher's Information may be expressed as

$$I(f) = \int_{-\infty}^{\infty} \left(\frac{f'(x)}{f(x)}\right)^2 f(x)dx \tag{1.17}$$

A brief outline of the rest of this dissertation will now be presented. Chapter II focuses on impulsive noise. First, the physical sources and properties of man-made and naturally occurring impulsive noise are discussed. Several models for impulsive noise are then presented. Well known empirical models (mathematical constructs designed to fit the data) and physical models (which are directly related to the underlying physical mechanisms) are described. Due to the complexity of these models and the need for simple density functions to approximate the first order characteristics of impulsive noise, three density systems are considered. They are: a generalized Gaussian noise, the Johnson S_u System, and a mixture model. Graphs of the density functions, their optimum and locally optimum non-linearities, and values for Fisher's Information are given for a range of values for the parameters of these systems. The chapter ends with a discussion of techniques which may be used for the detection of signals in an impulsive environment.

Frequently in detection problems, the statistical characteristics of the noise are not well established. In many cases, mini-max detectors may be used to achieve robustness. Chapter III presents results which facilitate the design of these detectors. First, the asymptotic performance (measured by the level and power of the test) of optimum detectors is related to the value of Fisher's Information for the underlying noise distribution. The density with minimum Fisher's Information over a given class of densities is seen to offer the worst environment in which to detect signals. This density, along with its locally optimum detector, is shown to form a saddlepoint. Several classes of densities are discussed, and conditions are established for the mini-max detectors.

Chapter IV presents an adaptive detection scheme. The Johnson S_u system of densities is used as a noise model. Quantiles are used to estimate the tail behavior of the first order density function of the noise. This estimate, along with an estimate of the variance, determines the detector's structure. Simulation results are presented which indicate that this detector will perform well over a large variety of background noises.

This dissertation ends with a brief summary and some suggestions for future research. An attempt has been made to keep all symbols used consistent throughout the text. Those symbols which were not presented in this introduction are presented as needed in the chapters. Chapters II, III, and IV are independent and may be read in any order.

References - Chapter I

1. H. M Hall, "A New Model for Impulsive Phenomena: Applications to Atmospheric-Noise Communication Channels," Stanford Electronics Laboratory, Stanford University, Stanford, CA, Tech. Rep. 3412-8 and 7050-7, August 1966.

2. E. N. Skomal, "Distribution and Frequency Dependence of Unintentionally Generated Man-Made VHF/UHF Noise in Metropolitan Areas," IEEE Trans. on Electromagnetic Compatibility, Vol. EMC-7, no. 4, pp. 420-427, Dec. 1965.

3. L. Batchelder, "Sonics in the Sea," Proc. IEEE, Vol. 53, no.10, pp. 1310-1319, October 1965.

4. A. D. Watt and E. L. Maxwell, "Measured Statistical Characteristics of VLF Atmospheric Radio Noise," Proc. IRE, Vol. 45, pp. 55-62, January 1957.

5. P. J. Bickel and K. A. Doksum, Mathematical Statistics, Holden-Day, Inc., San Francisco, 1977.

6. J. Capon, "On the Asymptotic Efficiency of Locally Optimum Detectors," IRE Trans. Infor. Theory, Vol. IT-7, no. 2, pp. 67-71, April 1961.

7. J. H. Miller and J. B. Thomas, "Detectors for Discrete-Time Signals in Non-Gaussian Noise," IEEE Trans. Infor. Theory, Vol. IT-18,no. 2, pp. 241-250, March 1972.

8. E. J. G. Pitman, Some Basic Theory for Statistical Inference, Chapman and Hall Ltd., London, 1979.

Chapter II

IMPULSIVE NOISE: CHARACTERISTICS, SOURCES, AND MODELS

This chapter serves as an introduction to impulsive noise. Section 1 contains a brief discussion of the characteristics and sources of impulsive noise. Section 2 presents several of the more common impulsive noise models. Empirical models designed to fit first order statistical data, and physical models directly related to the underlying physical mechanisms, are described. Most of these models are quite complex and do not lend themselves easily to detection problems. For this reason, Section 3 presents three density function systems which can be used to approximate the first order characteristics of impulsive noise. Section 4 ends the chapter with a discussion of techniques which may be used for the detection of signals in an impulsive environment.

Section 1: Characteristics and Sources

Impulsive noise can be characterized by a relatively small number of random, high amplitude bursts. Unlike either thermal noise or high density shot noise, each of which consists of a large number of sources contributing small disturbances and thus

satisfies the conditions of the Central Limit Theorem, impulsive noise typically consists of relatively few bursty sources. The Central Limit Theorem is not applicable, and the noise is distinctly non-Gaussian. The first order density functions usually have much heavier tails than that of the Gaussian distribution. Impulsive noise can rarely, if ever, be considered white. Samples are not independent. However, due to the complexities of studying n^{th} order non-Gaussian distributions, most research has centered on describing the first order characteristics. The independence assumption is then necessary to describe high order behavior.

Natural Noise Souces

Impulsive noise sources can be grouped into two categories: naturally occurring and man-made. Perhaps the naturally occurring impulsive noise which has captured the most attention in the literature [1,2,3,4] is atmospheric noise. It adversely affects communications systems from 1 kHz to 30 MHz. The total atmospheric noise measured at a given location can usually be treated as the sum of many individual lightning discharges modified by the appropriate propagation path. Due to the fact that at very low frequencies signals can propagate several thousands of miles, even extremely distant storms can raise havoc with radio communications. The following description of lightning discharges is due essentially to Watt and Maxwell [1].

A lightning discharge consists of both a predischarge and a main discharge. The predischarge consists of a series of short leader strokes, 30 to 200 feet in length, which attempt to establish an ionized path from the cloud to the ground. Each leader stroke is a current pulse of approximately 300 amperes and lasts about one microsecond. A new stroke occurs every 25 to 100 microseconds. This process forms what is termed the predischarge or fine structure of the lightning stroke. It typically lasts about one millisecond. According to Hall [4], the energy radiated by the predischarge has a 3 dB bandwidth of about 40 kHz with a maximum at approximately 30 kHz. The spectrum falls off as 1/f.

Once a leader reaches the ground, the main or return stroke follows the ionized path from the ground to the cloud. It consists of a 20,000 ampere current pulse and lasts about 100 to 200 microseconds. The return stroke accounts for about 95 percent of the total energy radiated by the lightning discharge. Often the main discharge is followed by one or more additional strokes which follow the same ionized path. The main discharge radiates power at a much lower frequency than the predischarge. The center frequency is about 10 kHz and the upper 3 dB point is approximately 15 kHz [4]. Beyond this, the spectrum decreases as $1/f^2$.

The majority of the power in VLF (3-30 kHz) and LF (30-300 kHz) atmospheric noise comes from the powerful return strokes. However, since the return strokes' power spectra decrease as $1/f^2$, predischarges contribute the majority of the atmospheric noise at

MF (300-3000 kHz) and HF (3-30 MHz). For still higher frequencies, atmospheric noise caused by lightning can usually be ignored.

Due to the many leader strokes of the predischarge and the occurrence of multiple return strokes, the received noise pulses may not be considered independent. However, neglecting the continuously changing effects of weather and propagation paths, these clusters of noise pulses may usually be considered independent of one another. In temperate climates a low frequency receiver will detect an average of one stroke every second [4]. In the tropics the average rate may be as high as 100 per second [4].

The sonar channel is also affected by naturally occurring impulsive noise. The major contributor in shallow water is marine life. Many species can produce underwater sounds which can mask desired signals [5,6]. Among the more important noise makers are snapping shrimp and croakers. Snapping shrimp, usually found in shallow tropical or subtropical water, make a crackling sound by repeatedly snapping closed their claws. Croakers, found predominantly in bays on the east coast of the United States, produce repeated bursts of tapping sounds. In areas with a large croaker population, the sound can resemble a continuous roar. Porpoise barks and gobbles, sea robin squawks, and toadfish "boops" may be among other additions to the chorus. These bursty sources of short duration sounds can yield a constantly changing, highly impulsive noise environment.

In polar regions, ice cracking and floe movement are the major contributors to acoustic impulsive noise. Milne and Ganton

[7] have measured acoustic noise, including amplitude distributions, under arctic sea ice. When the ice pack is solid, thermal stresses cause surface cracks which can be heard by a hydrophone below. At low frequencies (below 1 kHz) the noise is found to be impulsive. For frequencies above 1 kHz a more Gaussian noise has been recorded. They hypothesize that the low frequency noise is caused by medium to large thermal cracks in the surface ice. Higher frequency noise appears to be caused by numerous smaller thermal cracks located near irregularities on the ice surface. During the summer months many ice packs break up. Acoustic noise then results from the motion of the ice and collisions between ice floes. Under all but a relatively small range of frequencies, the summer ice noise is found to be non-Gaussian.

Man-made Noise Sources

Middleton [8] divides man-made impulsive noise sources on the radio channel into two categories: narrowband, for sources whose noise spectra are narrower than the bandpass of most receivers, and broadband. Automobile ignition noise, generally considered to be the most important VHF/UHF noise source in metropolitan areas [9], can be classified as broadband. It consists of random, very narrow noise pulses. Neon and fluorescent lights also contribute broadband noise [8,9]. The random impulses are less intense than those from automobile ignitions and can usually be considered a local disturbance. Other broadband impulsive radio noise sources are arc welders,

oil burner ignitions and other ignition noises. Narrowband noise includes emissions from high voltage transmission lines and generating stations, electrical motors, and various other electrical devices. Interference from other communications and radar systems can also be considered as narrowband impulsive noise.

The sonar channel is also affected by man-made noise [5,6]. Ships contribute appreciable acoustical noise in harbors and shipping lanes. Also, industrial plants along the coastline can be heard for a considerable distance offshore.

Telephone circuits must contend with switching noise. It is often modeled as an impulsive process [10,11]. The noise is broadband. Since the impulses usually occur in bursts, errors in digital transmissions are often bunched together.

Section 2: Impulsive Noise Models

Impulsive noise models are often grouped into two categories: empirical models and physical models. Empirical models are mathematical constructs which attempt to fit measured statistical data. Little or no attempt is made to relate the model to the physical mechanism of the process. Physical models, on the other hand, attempt to describe the entire noise process. Most consider the received noise as a sum of filtered impulses. Due to the mathematical difficulty involved with studying n^{th} order distributions, most research has concentrated on first order characteristics. The main

emphasis in the literature is on the modeling of the probability distribution of the narrowband noise envelope. Most work has centered on atmospheric noise.

Empirical Models

In most situations a received noise process, after having passed through the bandpass filter of the receiver, can be considered as a narrowband signal

$$f(t) = x(t)\cos[\omega t + \phi(t)] \tag{2.1}$$

When f(t) is a Gaussian process, samples of the envelope x(t) are random variables having a Rayleigh distribution. The Rayleigh distribution is given, in standard form, by

$$g(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} \quad , x>0 \tag{2.2}$$

$$= 0 \quad , x\leq 0$$

The mean is $0.5\sqrt{2\pi}\,\sigma$. The variance equals $(2-\pi/2)\sigma^2$.

It has been found [12] that the Rayleigh distribution is a close approximation to the low-amplitude, high-probability portion of the atmospheric noise envelope distribution. This is because the low-amplitude portion of the distribution is caused by the superposition of many overlapping small pulses which, according to the Central Limit Theorem, tend to behave like a Gaussian process. However, atmospheric noise, and

impulsive noise in general, has far more large amplitude spikes than predicted by the Rayleigh distribution. For this reason, the log-normal distribution was proposed [13]. Its density in standard form is

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}\ \frac{1}{x}\ e^{-\frac{1}{2\sigma^2}(\log x - \mu)^2} \qquad ,x>0 \qquad (2.3)$$

$$= 0 \qquad ,x\leq 0$$

The mean is $e^{\mu + \sigma^2/2}$. The variance is given by $e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$. Figure 2.1 displays both the Rayleigh and log-normal distributions. The parameters are chosen such that both densities have the same mean ($0.5\sqrt{2\pi} \approx 1.253$) and variance ($2-\pi/2 \approx 0.429$). The much heavier tail of the log-normal distribution can be seen in Fig. 2.2, which is an enlargement of the tail area of Fig. 2.1. Since tail behavior is very important in impulsive noise modeling, a more accurate method of display is required.

Impulsive noise data is often plotted on Rayleigh graph paper. Rayleigh coordinates are chosen such that the Rayleigh distribution plots as a straight line. For the cummulative distribution function $F(x) = \text{Prob}(x < t)$, one would plot $20\log(x/x_{rms})$ on the vertical axis and $-\log(-\ln[1 - F(x)])$ on the horizontal axis. Fig. 2.3 displays both the Rayleigh and log-normal distributions using Rayleigh coordinates. The heavier tail of the log-normal is now more readily apparent (left hand side of the graph).

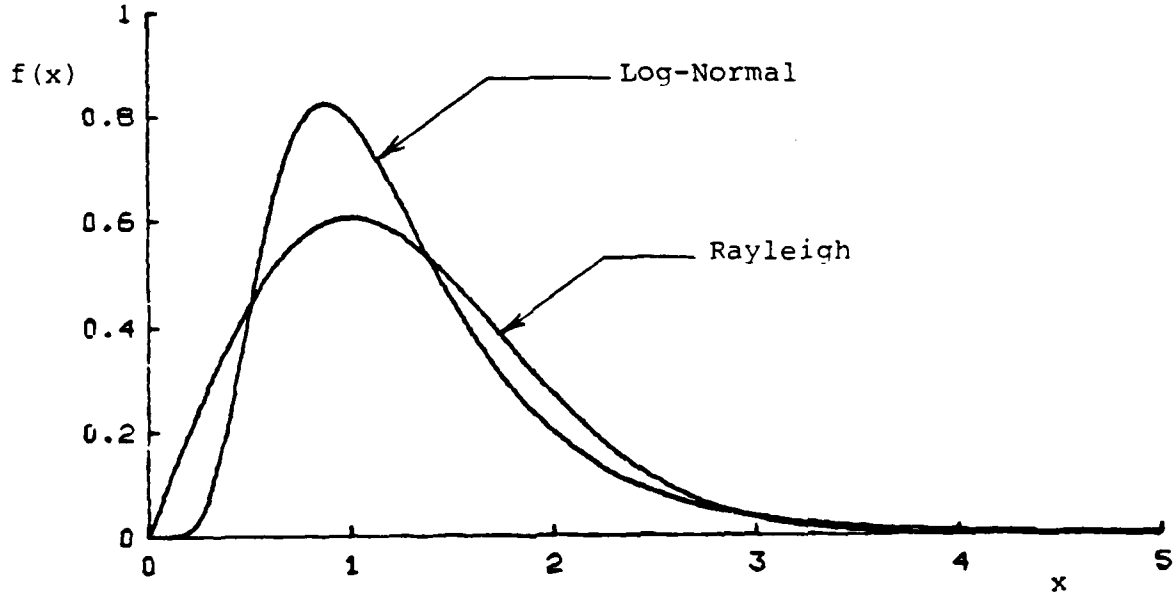


Fig. 2.1 - Rayleigh and Log-Normal Densities

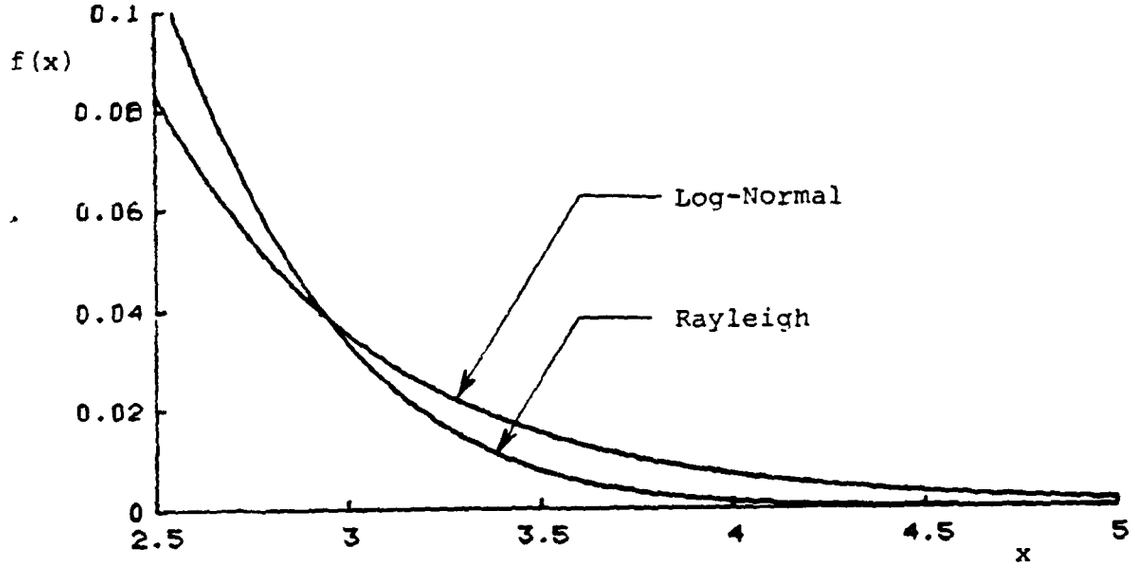


Fig. 2.2 - Tails of Rayleigh and Log-Normal Densities

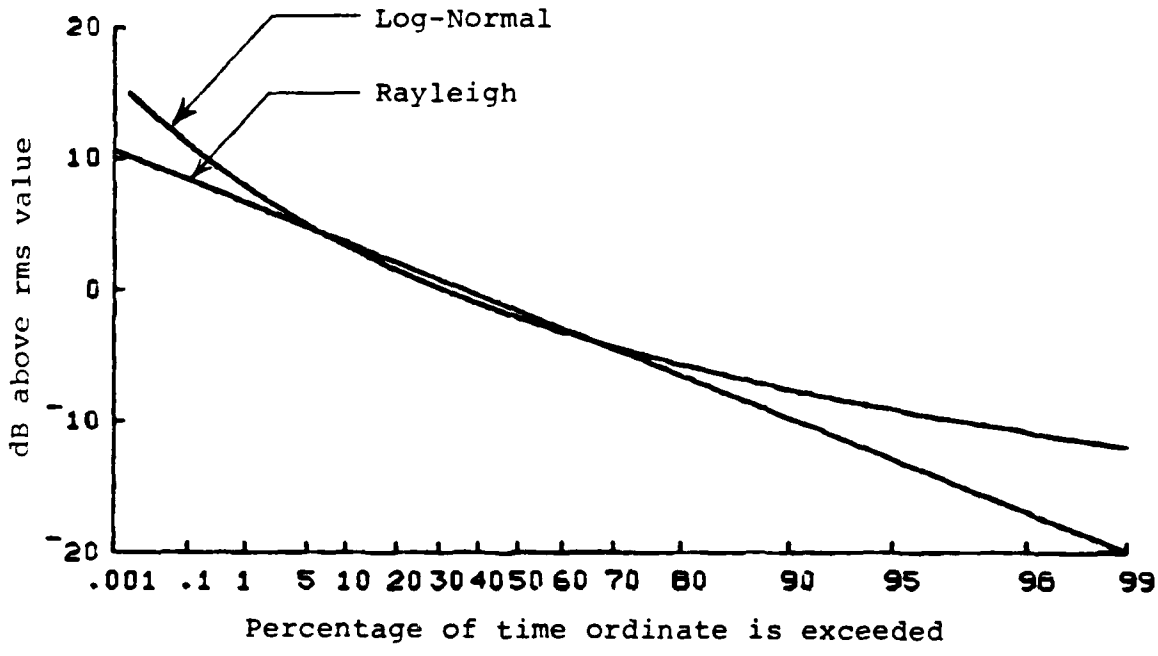


Fig. 2.3 - Rayleigh and Log-Normal Distributions
Rayleigh Coordinates

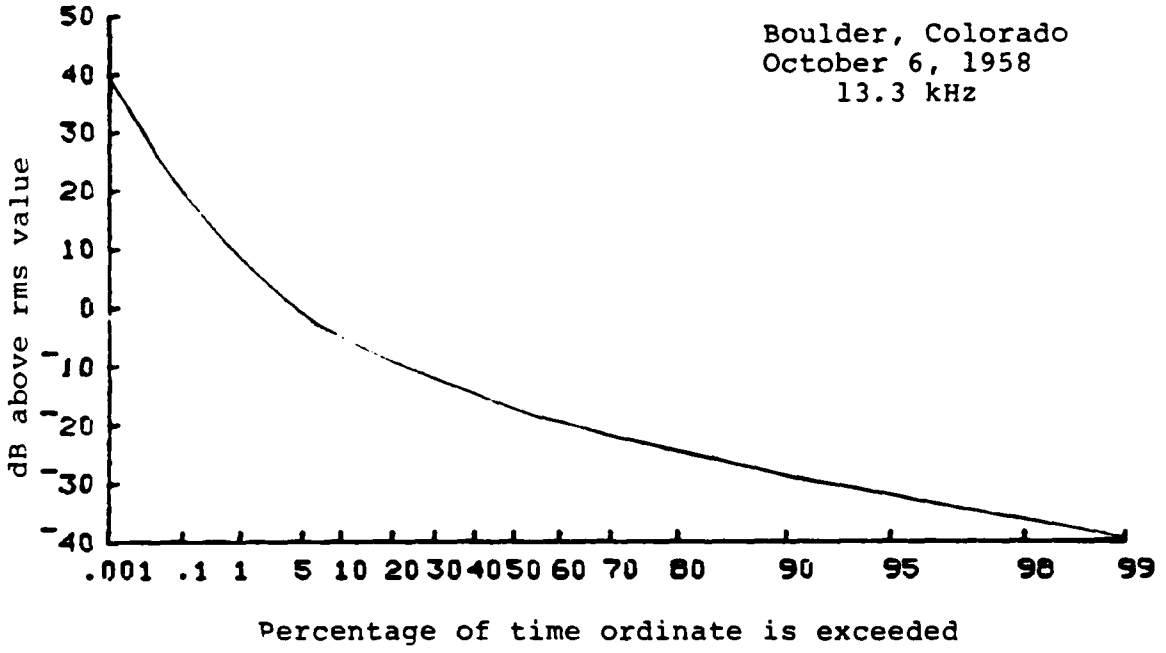


Fig. 2.4 - Measured Amplitude Probability Distribution of Atmospheric Radio Noise (from Critchlow [12])

Fig 2.4 shows atmospheric noise data measured at Boulder, Colorado [12]. The shape of the curve is typical for broadband impulsive noise. Neither the Rayleigh nor the log-normal distribution provides an acceptable fit over the entire range of values.

A graphical technique to derive the envelope probability distribution function for atmospheric noise, given measurements of three moments, was developed at the National Bureau of Standards [12]. These three moments: the average noise power, the average envelope voltage, and the average logarithm of the envelope voltage, were continuously recorded during the International Geophysical Year (1958) at sixteen stations throughout the world. The graphical method was developed to utilize this information to accurately model atmospheric noise.

Analyzing the data, it was noticed that the atmospheric noise envelope probability distribution, when plotted using Rayleigh coordinates, could be approximated by a three section curve. The low-amplitude, high-probability portion corresponds to the Rayleigh line. The high-amplitude, low-probability portion can be approximated by a straight line of somewhat steeper slope. It can be shown [14] that all functions of the form $F(x) = 1 - \exp(-kx^m)$ (and only these functions) will map on Rayleigh coordinates as straight lines. This family of curves, which includes the Rayleigh distribution, is often called the power-Rayleigh distribution. The middle amplitude portion of the curve was then approximated by the arc of a circle tangent to the two lines. The center of this circle lies on the bisector of the angle formed by the Rayleigh and power-Rayleigh lines.

Four parameters are needed to specify this approximation to the true distribution:

(1) a point on the Rayleigh line

(2) a point on the power-Rayleigh line

(3) the slope of the power-Rayleigh line

and (4) the radius of the circle

Empirical relationships were developed [12] to obtain these four parameters from the three measured moments. Thus, by measuring the three moments, a function can be obtained which is a fairly accurate approximation to the envelope probability distribution function of the noise. A method was later developed [15] to obtain the distribution function for any bandwidth by transforming the distribution function obtained for the bandwidth of the measurements. These graphical methods, although fairly accurate, were not given any theoretical justification. Even more importantly, the procedures are of limited use in the analytical treatment of optimum detectors and estimators.

Many other distributions have been suggested to model impulsive noise (see [8] for a concise list). Most are limited in application and not entirely successful. None of the empirical models give a true picture of the entire interference process.

Physical Models

Most physical models attempt to describe the entire interference process by modeling the received noise as a sum of filtered impulses:

$$x(t) = \sum_{i=1}^{N} p(a_i;\ t-t_i) \tag{2.4}$$

The $\{a_i\}_1^N$ are random variables which describe the pulse amplitudes, $p(\cdot)$ describes the pulse shape, and the random variables $\{t_i\}_1^N$ are the occurrence times. Different assumptions for these variables can lead to quite different models.

Furutsu and Ishida [16] were interested in modeling the HF radio noise which is predominantly caused by the fine structure of the lightning discharge. They viewed each predischarge as a noise packet. They assumed Poisson occurrence times for these packets. Within each packet they modeled the fine structure as another independent Poisson process. This model, consisting of Poisson noise packets occurring in a Poisson manner, is often called Poisson-Poisson noise.

In order to study the narrowband noise envelope, Furutsu and Ishida assume the following filtered impulse model:

$$x(t) = \sum_{i=1}^{N} \sum_{j=1}^{N'} r(t,\ t_i,\ t'_{ij},\ a_{ij})\cos[\omega(t-t'_{ij})-\phi_{ij}] \tag{2.5}$$

where t = the time of observation

t_i = time of occurrence of i^{th} pulse packet

t'_{ij} = time of occurrence of the j^{th} pulse within the i^{th} packet

ω = receiver center frequency

ϕ_{ij} = phase of the j^{th} pulse within the i^{th} packet

They assume the phase is uniformly distributed over $(0,2\pi)$. The packet occurrence times are assumed to be uniformly distributed over the interval of observation $(0,T)$ and the random pulses are assumed uniformly distributed over the length of each packet. The probability that N' pulses occur in time $(0,t)$, within a packet, is given by

$$P_t'(N') = \frac{(\nu' t)^{N'}}{N'!} e^{-\nu' t} \tag{2.6}$$

The probability that N packets occur in $(0,T)$ is given by

$$P_T(N) = \frac{(\nu T)^{N}}{N!} e^{-\nu T} \tag{2.7}$$

For clarity, let $\tau = t - t_i$ = the length of time from the start of the i^{th} packet to the time of the observation, $a = a_i$, and $t' = t'_{ij}$. Now assume

$$r(t,t_i,t'_{ij},a_{ij}) = r(\tau,t',a) = ae^{-\beta(\tau-t')} e^{-\alpha t'} \tag{2.8}$$

Here $e^{-\alpha t'}$ is the receiver impulse response and $ae^{-\beta(\tau-t')}$ is the amplitude of the impulse within the packet. Note that, although 'a' is a random variable, the impulses will tend to be larger at the start of a packet and then will tend to decrease in magnitude. Also, 'a' is assumed distributed according to w(a), some function which is chosen considering the spatial distribution and strength of the impulse sources.

Using the above assumptions, Furutsu and Ishida calculated the first order density function for the noise envelope. Their results are

$$f(E) = \int_0^\infty E\lambda \, J_o[E\lambda] \, \psi(\lambda) \, d\lambda \tag{2.9}$$

Here $J_o[\cdot]$ is the zero order Bessel function (1^{st} kind) and $\psi(\lambda)$ is the characteristic function,

$$\psi(\lambda) = \exp \{ \nu \int_0^T [\psi'(\lambda) - 1] \, dt \} \tag{2.10}$$

where,

$$\psi'(\lambda) = \exp \{ \nu' \int_0^\tau \int_0^\infty w(a) \, (J_o[\lambda r(\tau,t',a)] - 1) \, da \, dt' \} \tag{2.11}$$

Upon substitution of $r(\tau,t',a)$ and $w(a)$ the integrals become quite intractable. Simplifying assumptions lead to solutions for special cases. These solutions have shown fairly good agreement with measurements. However, since the density functions cannot be put in closed form in general, this model is of only limited use in the analytical treatment of detection and estimation problems.

Beckmann [14] proposed a different model for atmospheric noise. He assumed individual atmospherics could be represented by pulses, whose envelopes could be written as

$$u_k = E_k\, e^{-(t-t_k)/a} \quad , \; t \geq t_k \qquad (2.12)$$

$$= E_k\, e^{(t-t_k)/b} \quad , \; t < t_k$$

The rate of decay for the spikes was assumed to be much slower than the rate of upsurge ($b<<a$). For convenience, define $c = (a+b)/2$, the average time constant. Beckmann chose the peak amplitude E_k to be log-normally distributed. He justified this as follows. Divide the propagation path for a given atmospheric into a large number of sections. Assume the attenuation for each section is independent of the rest and that no one section's attenuation will dominate. Then the total attenuation (expressed in dB) will be normally distributed (mean μ, variance σ^2). Since the log of the peak amplitude (E_k) is proportional to the total attenuation, the peak amplitude is log-normally distributed.

Using Poisson occurance times (with ν = the mean number of pulses per unit time) and considering the phase to be uniformly distributed, the values for all the pulses, both rising and decaying, are summed. In the low density case ($\nu c<<1$), the nearest pulse will dominate the others. The rest can be approximated by a single pulse with a Rayleigh envelope. Thus the received noise can be modeled as the sum of a pulse with a log-normally distributed envelope and a pulse with a Rayleigh envelope. Beckmann calculates the distribution function of the total noise envelope as

$$f(E) = \frac{2E}{M} \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{y} \exp\left[-\frac{(\ln y - \mu)^2}{2\sigma^2} - \frac{E^2 + y^2}{M}\right] I_o\left(\frac{2yE}{M}\right) dy \qquad (2.13)$$

where $M = \nu^2 c^2 e^{2(\sigma^2+\mu)} \ln(1/\nu c)$ is approximately equal to the noise power and $I_o(\cdot)$ is the 0^{th} order modified Bessel function. This distribution is nearly Rayleigh for small E and nearly log-normal for large values of E. The four parameters of the density ν, c, σ and μ can be directly related to the properties of the noise. Both ν, the mean number of discharges per unit time, and c, the average time constant, depend only on the properties of the discharges. The parameter σ, the standard deviation of the total attenuation of the pulses, and μ, the mean value of the total attenuation, are determined almost entirely by the properties of the propagation path. Also, when plotted on Rayleigh coordinates, νc determines the position of the Rayleigh line (low-amplitude, high-probability line), σ determines the log-normal curve (high-amplitude, low-probability), and ν, c, σ and μ together determine the rms value. Thus, knowing these four parameters, a density can be sketched by connecting these two functions with a circular arc as was done by Crichlow, et al. [12].

Although Beckmann's model is closely related to the statistical parameters of the noise process, the integral of Eq. (2.13) is rather cumbersome. A general solution can only be found for very small or very large envelope levels. Hence, Beckmann provides a theoretical justification for the use of the log-normal distribution for strong atmospherics, but he

doesn't provide a density function which can represent the noise over the entire range of interest. Giordano [17] and Middleton [8,18,19,20] have both developed filtered impulse models which are much more widely applicable. Giordano's model is similar to those previously described and will not be treated further. Middleton's model is a more complete and a more general model than those presented above.

Middleton represents the received noise process as

$$y(t) = x(t) + n(t) \tag{2.14}$$

where $x(t)$ is the impulsive component and $n(t)$ is a Gaussian background (variance σ_G^2). He assumes

$$x(t) = \sum_j U_j(t, \underline{\theta}) \tag{2.15}$$

where $U_j(t, \underline{\theta})$ is the waveform emitted by the j^{th} source after having passed though the receiver. All received noise pulses are assumed to have the same basic waveform with random variations in scale and structure represented by the random parameter $\underline{\theta}$. Middleton assumes the sources emit independently in time according to a Poisson distribution with mean rate ν. He denotes the mean duration of a pulse by T. The source locations are assumed Poisson distributed in a space Λ with coordinates $\underline{\lambda}$.

Middleton uses characteristic functions to derive the density function for the noise. He considers two cases: narrowband (noise whose spectra are narrower than the passband of the receiver) and broadband (noise with spectra comparable to

or wider than the passband of the receiver). Note, however, that both cases produce narrowband noise in the receiver. That is, in both cases the noise can be modeled by its envelope and phase. For both the narrowband and broadband cases, the characteristic function for x(t) can be written,

$$\psi_x(u) = \exp[E\{A\, J_o(Bu) - A\}] \tag{2.16}$$

where $J_o(\cdot)$ is the 0^{th} order Bessel function (1^{st} kind), $B = B(t, \underline{\lambda}, \underline{\theta})$ is the envelope of $U(t, \underline{\theta})$ (which can be directly related to the physics of the noise generation and propagation), and $A = \sqrt{T}$ is the first basic parameter of the model. The parameter "A" is termed the "Impulsive Index." It measures the amount that the noise pulses overlap in time. For high levels of overlap (large A), the density will approach the Gaussian distribution. When the amount of overlap is small (small A), the resulting noise will be highly impulsive.

For narrowband noise, several approximations can be made in Eq.(2.16). The resulting characteristic function for x(t) is

$$\psi_x(u) \approx e^{-A} \sum_{m=0}^{\infty} \frac{A^m}{m!}\, e^{-\frac{m}{4} E\{B^2\}u^2} \tag{2.17}$$

The characteristic function for y(t) is then

$$\psi_y(u) = \psi_x(u) \cdot e^{-\frac{1}{2}\sigma_G{}^2 u^2} \tag{2.18}$$

$$\approx e^{-A} \sum_{m=0}^{\infty} \frac{A^m}{m!}\, e^{-\frac{m}{4} E\{B^2\}u^2 - \frac{1}{2}\sigma_G{}^2 u^2}$$

This function can now be inverted. Letting z be the normalized noise ($z = y/y_{rms}$), Middleton finds for the probability density function of the instantaneous amplitude in the narrowband case:

$$f(z) \approx e^{-A} \sum_{m=0}^{\infty} \frac{A^m}{m!\sqrt{2\pi}\sigma_m} e^{-\frac{z^2}{2\sigma_m^2}} \tag{2.19}$$

where

$$\sigma_m^2 = \frac{m/A + P}{1 + P} \tag{2.20}$$

and "P" is the ratio of the power in the Gaussian background to that in the impulsive component

$$P = \frac{\sigma_G^2}{1/2[A\ E\{B^2\}]} \tag{2.21}$$

It is the second basic parameter for the narrowband model. Note that Eq. (2.19) is just a weighted sum of Gaussian distributions with increasing variance σ_m^2. For the envelope of the received noise, the model yields

$$\text{Prob}\ [E>E_o] \approx e^{-A} \sum_{m=0}^{\infty} \frac{A^m}{m!} e^{\frac{-E_o^2}{\sigma_m^2}} \tag{2.22}$$

which is just a weighted sum of Rayleigh distributions.

In the broadband case the simplifying assumptions for Eq. (2.16) can no longer be used. The procedure becomes more difficult but is still tractable. One complication is that the resulting density has infinite variance for certain ranges of the parameters. Thus, to normalize the noise, the power in the Gaussian background alone is used: $z = y/n_{rms}$. The density can be written

$$f(z) \approx \frac{1}{\pi} e^{-A_\alpha} \sum_{m=0}^{\infty} \frac{(-1)^m}{m!} A_\alpha^m \Gamma\left(\frac{m\alpha+1}{2}\right) {}_1F_1\left(-\frac{m\alpha}{2}, \frac{1}{2}, z^2\right) \tag{2.23}$$

where ${}_1F_1(\cdot,\cdot,\cdot)$ is the confluent hypergeometric function [21], $\Gamma(\cdot)$ is the gamma function and α, A_α are the two basic parameters for the model. They are both related to the source distributions and propagation laws for the noise. The envelope in the broadband case is

$$\text{Prob}[E>E_o] \approx e^{-E_o^2} \left[1 - E_o^2 \sum_{m=1}^{\infty} \frac{(-1)^m}{m!} A_\alpha^m \Gamma\left(1 + \frac{m\alpha}{2}\right) \cdot \right. \tag{2.24}$$

$$\left. \cdot {}_1F_1\left(1 - \frac{m\alpha}{2}, 2, E_o^2\right)\right]$$

Middleton has presented evidence that Eqs. (2.22) and (2.24) are accurate representations for the envelope of impulsive noise. Figs. 2.5, 2.6, 2.7, and 2.8 are taken from [8]. Fig. 2.5 shows a comparison between Middleton's narrowband model (A=0.35, P= $0.5\text{x}10^{-3}$) and actual narrowband impulsive noise data. Fig. 2.6 is a similar display ($A=10^{-4}$, P=50) for narrowband electromagnetic noise from ore crushing machinery.

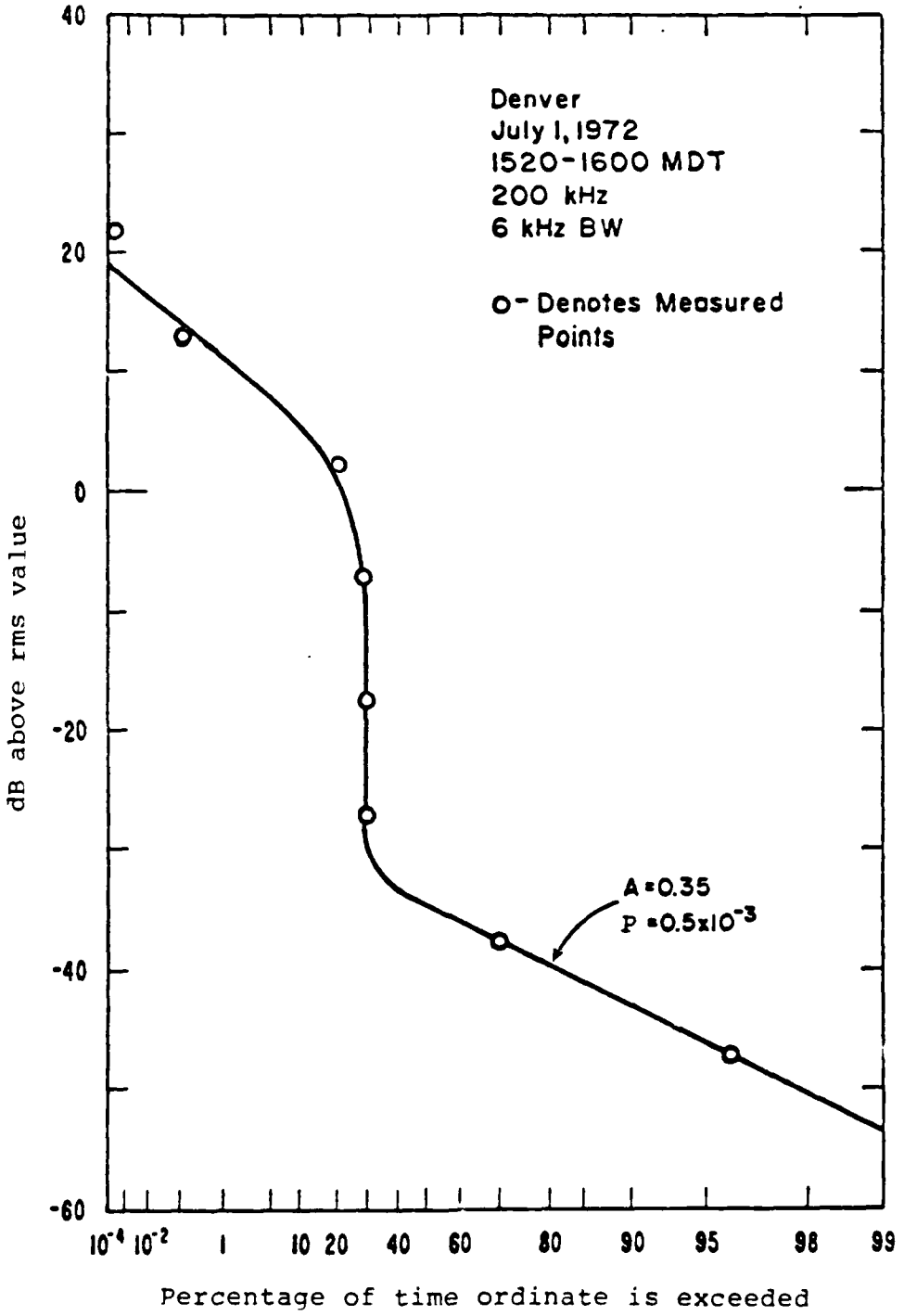


Fig. 2.5 - Comparison of a Measured Envelope Distribution with the Middleton Model - Narrowband Case (from Spaulding and Middleton [8])

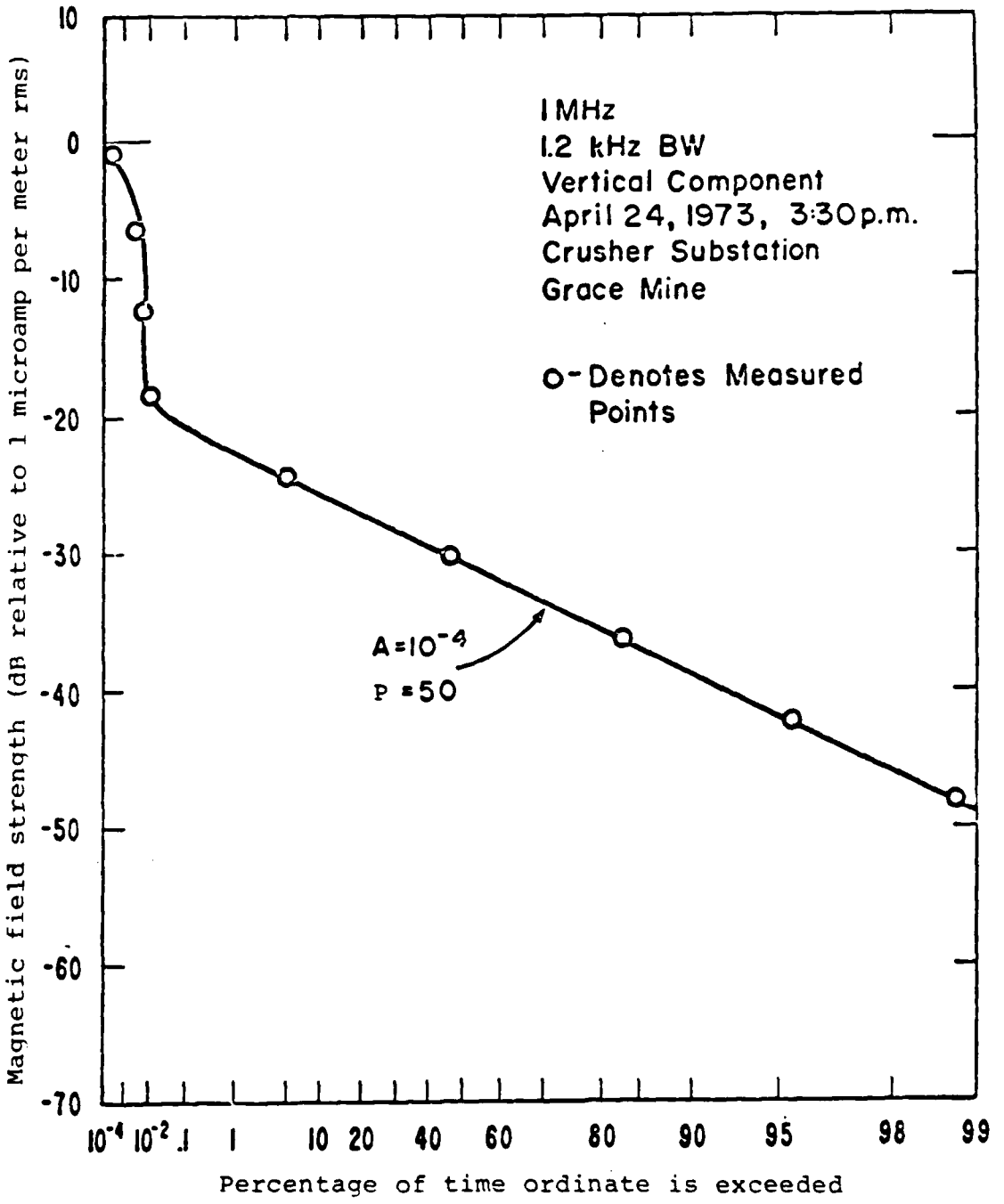


Fig. 2.6 - Comparison of a Measured Envelope Distribution with the Middleton Model - Narrowband Case (from Spaulding and Middleton [8])

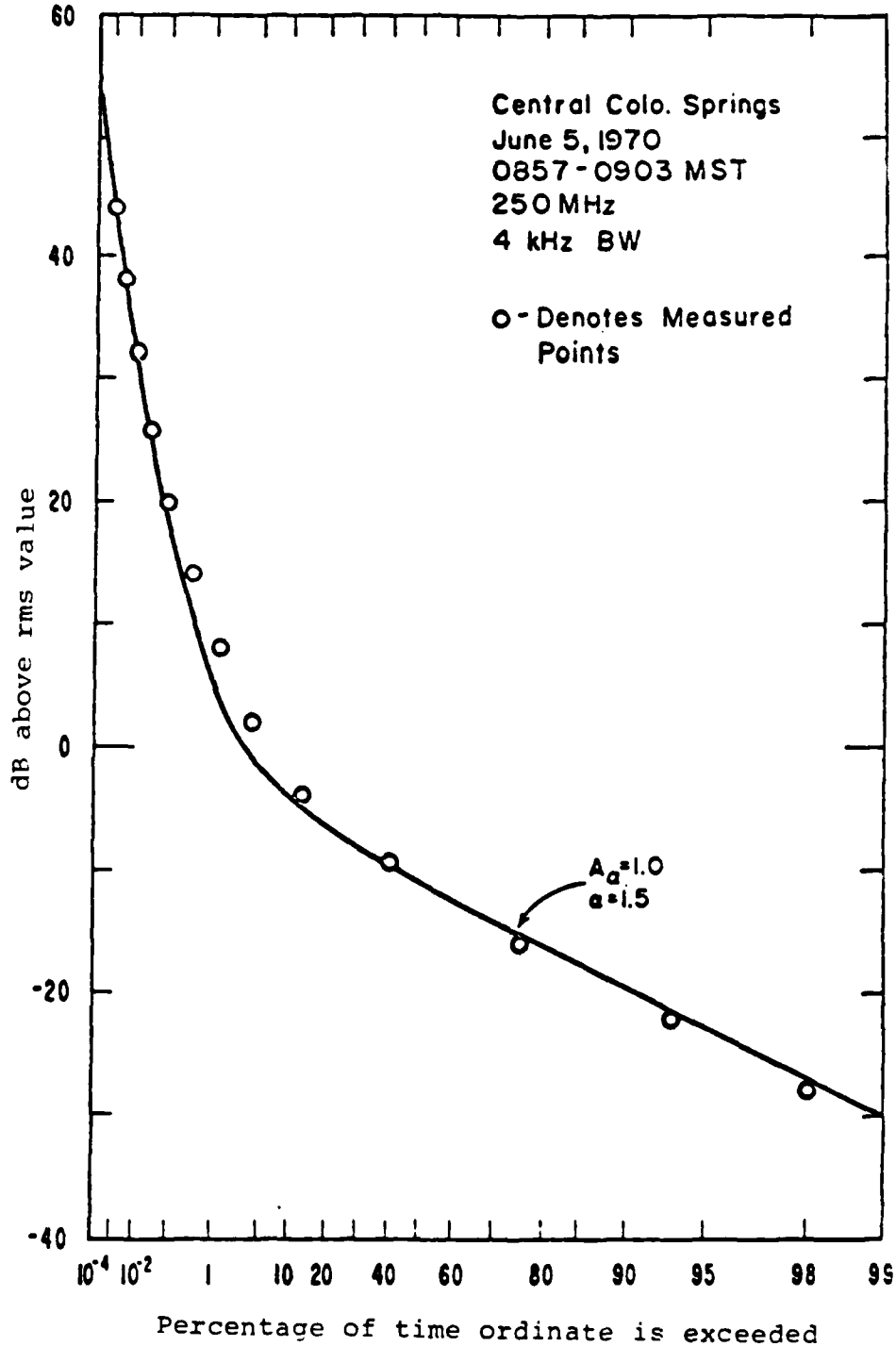


Fig. 2.7 - Comparison of a Measured Envelope Distribution with the Middleton Model - Broadband Noise (Man-Made Noise) (from Spaulding and Middleton [8])

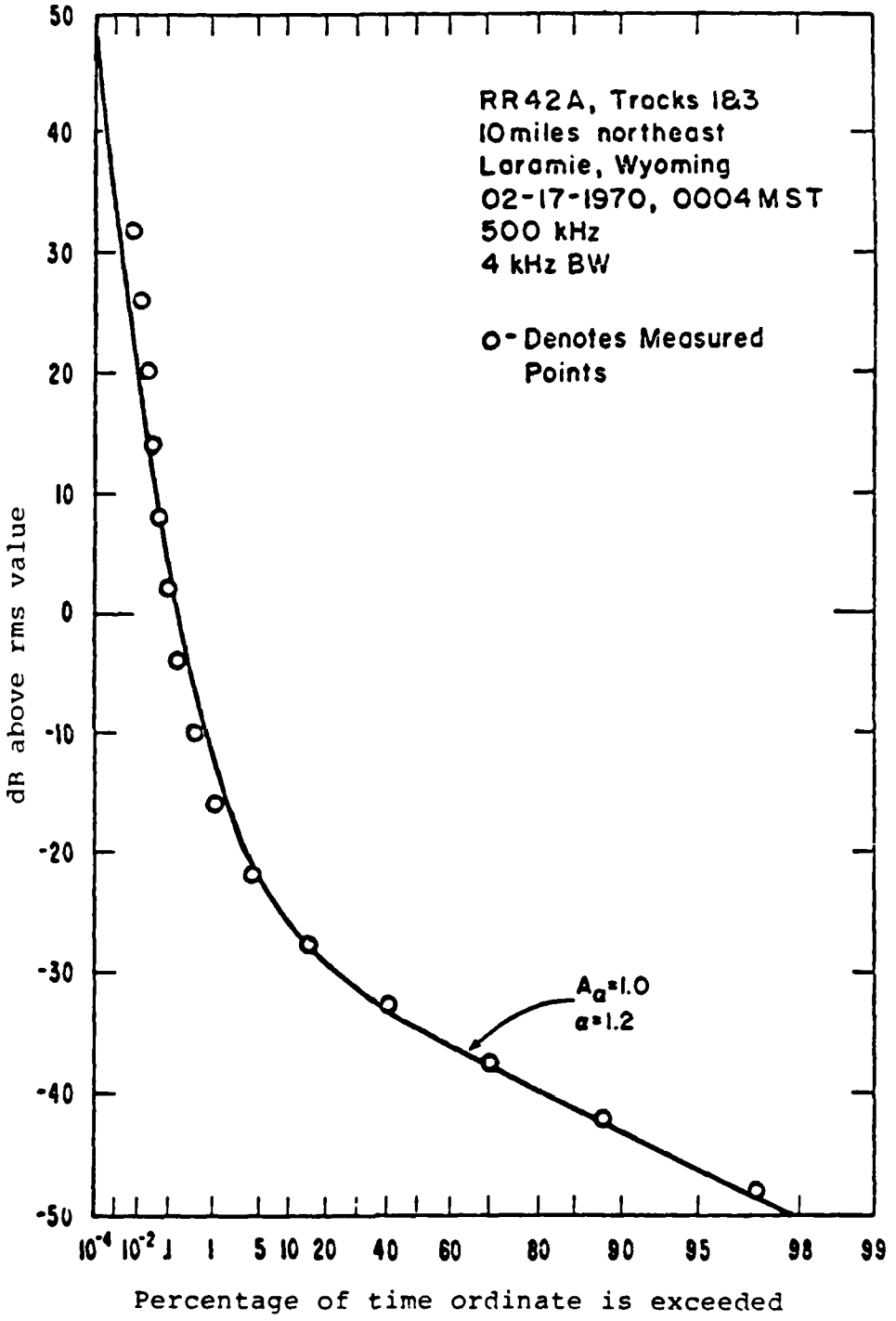


Fig. 2.8 - Comparison of a Measured Envelope Distribution with the Middleton Model - Broadband Case (Atmospheric Noise) (from Spaulding and Middleton [8])

In both cases agreement between the model and data is quite good. Figs. 2.7 and 2.8 compare Middleton's broadband model with measured impulsive noise data. Fig. 2.7 shows broadband man-made noise (primarily automobile ignition noise) and Fig. 2.8 displays atmospheric noise. Again close agreement between the model and measurements are found.

The most important aspect of the Middleton model which sets it apart from the rest is that the parameters for this model (A and P or α and A_α) can be determined explicitly from physical considerations. The model can be tested by the data rather than just fit to the data. The chief problem with this model is the assumption of independent noise pulses. This often is not valid. Unfortunately, as with the other models, the equations are intractable without this assumption.

Another physical model was proposed by Hall [4]. Although not as comprehensive as the Middleton model, it deserves attention because it is not of the filtered impulse type. Hall was interested in modeling VLF atmospheric noise. In order to achieve a large dynamic range, he proposed a model which considered the received noise process to be of the form

$$y(t) = a(t)n(t) \tag{2.25}$$

where a(t) is a slowly varying stationary random process and n(t) is an independent, narrowband Gaussian process. Hall selected a first order distribution for a(t) which was analytically tractable and which yielded good agreement with measured atmospheric noise data. His distribution for a(t) is:

$$f_a(a) = \frac{(m/2)^{m/2}}{\sigma^m \Gamma(m/2)} \frac{1}{|a|^{m+1}} e^{-\frac{m}{2a^2\sigma^2}} \tag{2.26}$$

where m and σ are two parameters to be chosen. Using the first order distribution for the Gaussian process,

$$f_n(n) = \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{n^2}{2\sigma_1^2}} \tag{2.27}$$

he calculated the first order probability density function for y(t) as

$$f_y(y) = \frac{\Gamma(\frac{\theta}{2})}{\Gamma(\frac{\theta-1}{2})} \frac{\gamma^{\theta-1}}{\sqrt{\pi}} \frac{1}{(y^2+\gamma^2)^{\theta/2}} \tag{2.28}$$

where

$$\gamma = m^{\frac{1}{2}} \frac{\sigma_1}{\sigma}$$

$$\theta = m + 1$$

For the special case $\sigma_1 = \sigma$, Eq. (2.28) reduces to the density for Student's "t". Hence, Hall named Eq. (2.28) the generalized "t" distribution.

As a check on his model, Hall calculated the first order distribution function for the envelope and phase of y(t). He obtained

$$f_E(E) = (\theta-1)\,\gamma^{\theta-1} \frac{E}{(E^2+\gamma^2)^{\frac{\theta+1}{2}}} \tag{2.29}$$

and a uniform phase. For θ in the range $2<\theta\leq 4$, Eq. (2.29) agrees well with VLF atmospheric noise data. The value $\theta=3$ is best; however, the density function [Eq. (2.28)] has infinite variance for $\theta=3$ and thus can not represent a real process.

Although Hall's model is simple enough to use in detection problems and is fairly accurate, it does not have a strong physical basis. The parameters θ and γ have no real physical meaning and must be selected to fit statistical data rather than physical conditions.

Section 3: Tractable First Order Non-Gaussian Noise Models

Although the models described in Section 2 are very detailed and many are quite accurate, they are all quite complicated. Most cannot be used in analytical treatments of detection problems. Those which can are cumbersome. There is a need for simple, first order families of densities which can be used to describe the first order statistics of non-Gaussian noise. Three possibilities are: a generalized Gaussian noise, the Johnson S_u System, and mixture models. These three systems are used throughout this dissertation to provide examples.

A Generalized Gaussian Noise

A generalized Gaussian noise [22], which provides a system of density functions whose tails can be made heavier or lighter than that of the Gaussian distribution, is given by

$$f(x) = \frac{c\eta(\sigma,c)}{2\Gamma(1/c)} e^{-[\eta(\sigma,c)|x|]^c} \qquad ,c>0 \tag{2.30}$$

where

$$\eta(\sigma,c) = \frac{1}{\sigma}\left[\frac{\Gamma(3/c)}{\Gamma(1/c)}\right]^{1/2}$$

The case c=2 is the familiar Gaussian distribution and c=1 yields the Laplace or double exponential distribution. The density functions are plotted in Fig. 2.9 for values of c in the region of most interest, c=0.5 to c=3.

The locally optimum nonlinearities for detecting a constant signal in this noise are given by

$$g_{loc}(x) = c[\eta(\sigma,c)]^c |x|^{c-1} \operatorname{sgn}(x) \qquad ,c>0 \tag{2.31}$$

They smoothly cover the transitions from a blanker, whose tail region decreases as 1/x, through the sign and linear detectors, to an expander. The locally optimum detectors corresponding to the densities of Fig. 2.9 are given in Fig. 2.10. The c=2 case

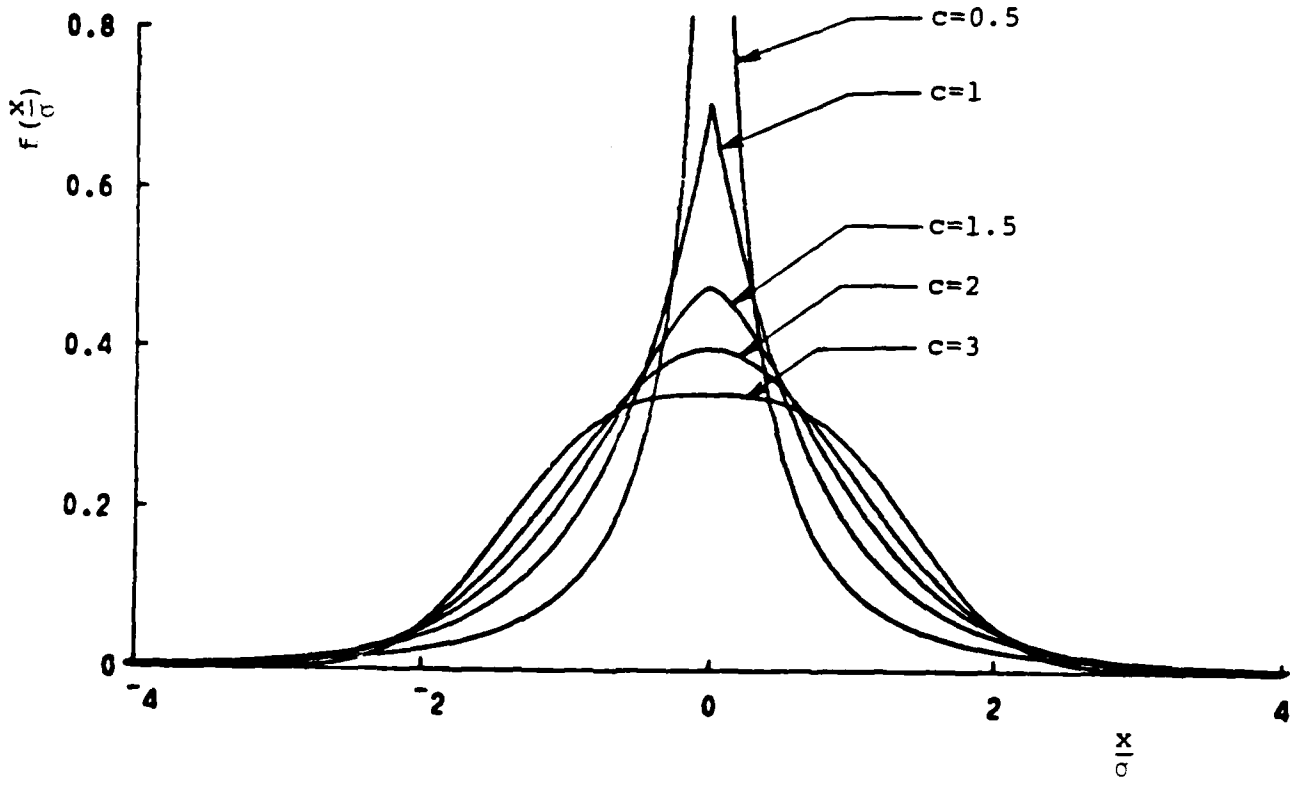


Fig. 2.9 - Generalized Gaussian Noise: Density Functions

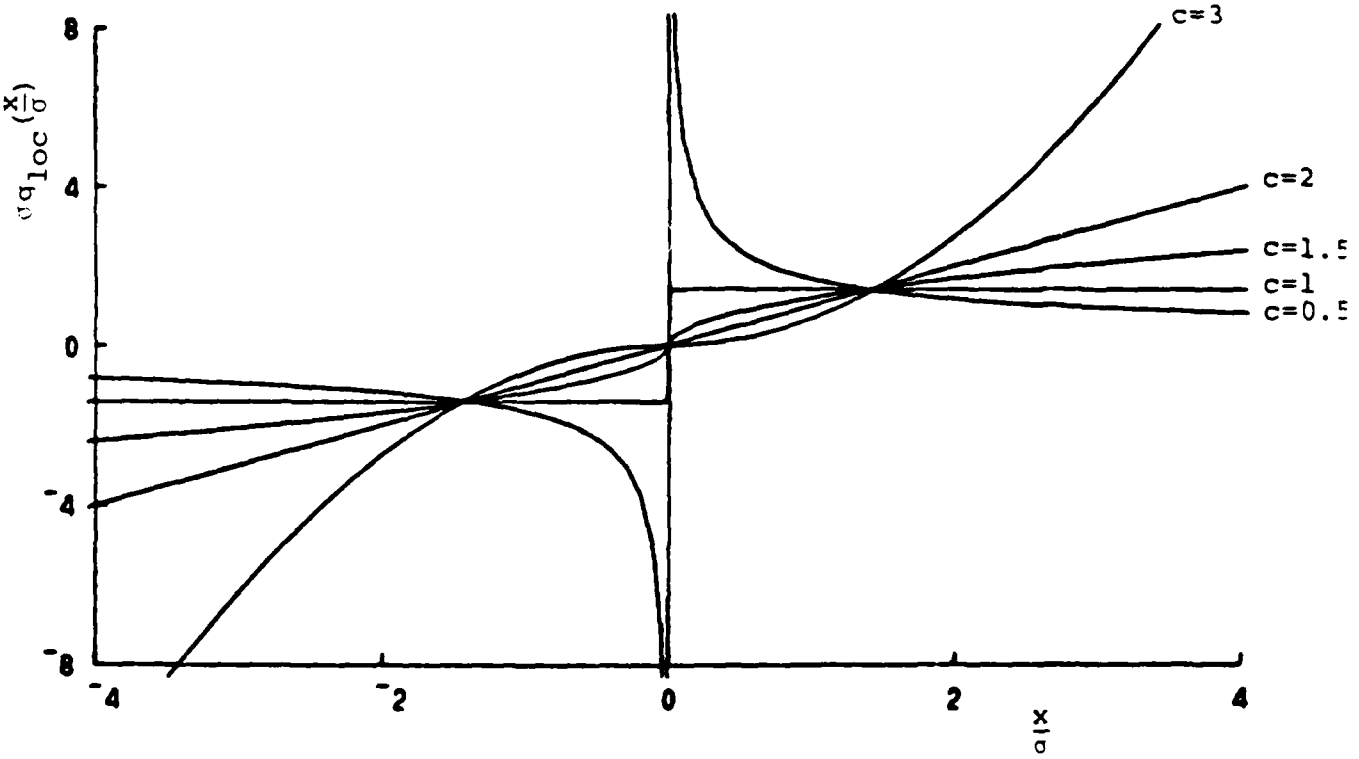


Fig. 2.10 - Generalized Gaussian Noise: Locally Optimum Nonlinearities

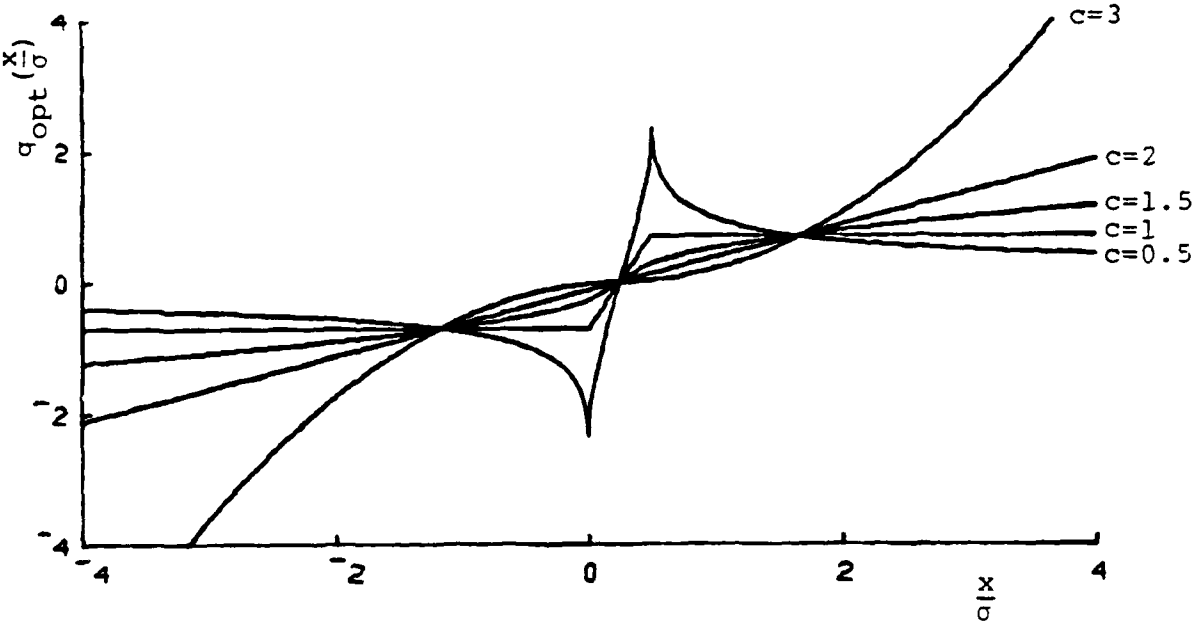


Fig. 2.11 - Generalized Gaussian Noise: Optimum Nonlinearities (signal strength = $\frac{\sigma}{2}$)

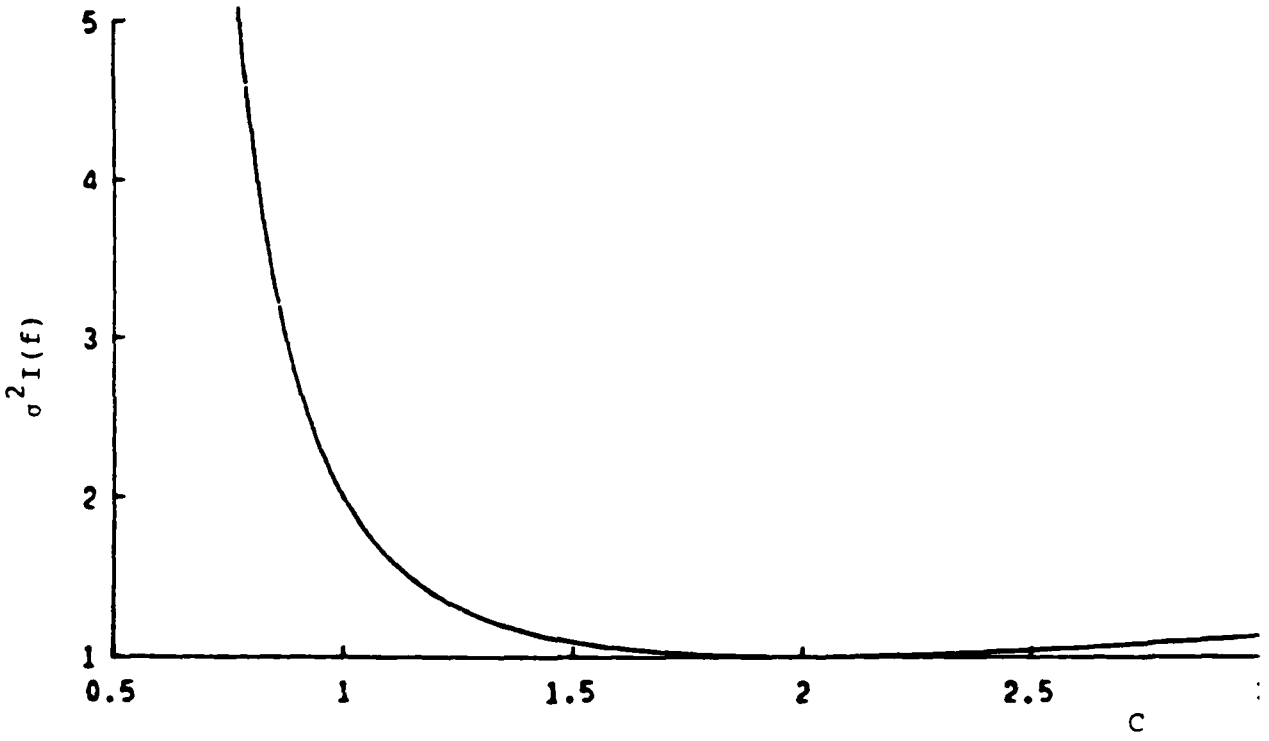


Fig. 2.12 - Generalized Gaussian Noise: Fisher's Information vs. C

is of course the linear detector and c=1 gives the hard limiter or sign detector. The Neyman-Pearson optimum nonlinearities for the constant signal case are

$$g_{opt}(x) = [\eta(\sigma,c)]^c \ (|x|^c - |x-\theta|^c) \qquad ,c>0 \tag{2.32}$$

These are plotted in Fig. 2.11 for the signal strength $\theta=\sigma/2$.

Fisher's Information can be calculated by substituting Eq. (2.30) into Eq. (1.17). After integrating [see Appendix 2.1] one obtains,

$$I(f) = \frac{c^2 \ \Gamma(\frac{3}{c}) \ \Gamma(2-\frac{1}{c})}{\sigma^2 \ [\Gamma(\frac{1}{c})]^2} \qquad ,c>1/2 \tag{2.33}$$

$$= \infty \qquad ,0<c\leq 1/2$$

This function is plotted in Fig 2.12. Since the Gaussian distribution minimizes Fisher's Information over all <u>fixed variance</u> distributions, the curve has a minimum at c=2. The peaked origin of the density function accounts for the greater information for location for c<2, and the greatly diminished tails account for the case c>2.

<u>The Johnson S_u System</u>

The Johnson S_u System [23] can be used to provide a system of heavy-tailed symmetric distributions. It is formed by a memoryless transformation of the Gaussian distribution,

$$x = a + \lambda \sinh\left(\frac{z - \gamma}{\delta}\right) \tag{2.34}$$

The resulting density functions can be written

$$f(x) = \frac{1}{\sqrt{2\pi}} \frac{\delta}{\lambda} \left[1 + \left(\frac{x-a}{\lambda}\right)^2\right]^{-1/2} e^{-1/2\{\gamma + \delta \sinh^{-1}\left(\frac{x-a}{\lambda}\right)\}^2} \tag{2.35}$$

The parameters a and γ, which affect location and symmetry are set equal to zero. The parameters δ and λ affect the shape and scale of the densities. The choice of

$$\lambda = \left[\frac{2\sigma^2}{e^{2/\delta^2} - 1}\right]^{1/2} \tag{2.36}$$

yields densities of common variance σ^2. Thus the densities can be specified by the two parameters σ^2 and δ. A few densities from the Johnson System are plotted in Fig. 2.13. Small values of δ correspond to heavy-tailed distributions. In the limit as $\delta \to \infty$, the densities approach the Gaussian distribution.

The locally optimum nonlinearities for the detection of constant signals are given by

$$g_{loc}(x) = \left[1 + \left(\frac{x}{\lambda}\right)^2\right]^{-1} \left(\frac{x}{\lambda^2}\right) + \left[1 + \left(\frac{x}{\lambda}\right)^2\right]^{-1/2} \frac{\delta^2}{\lambda} \sinh^{-1}\left(\frac{x}{\lambda}\right) \tag{2.37}$$

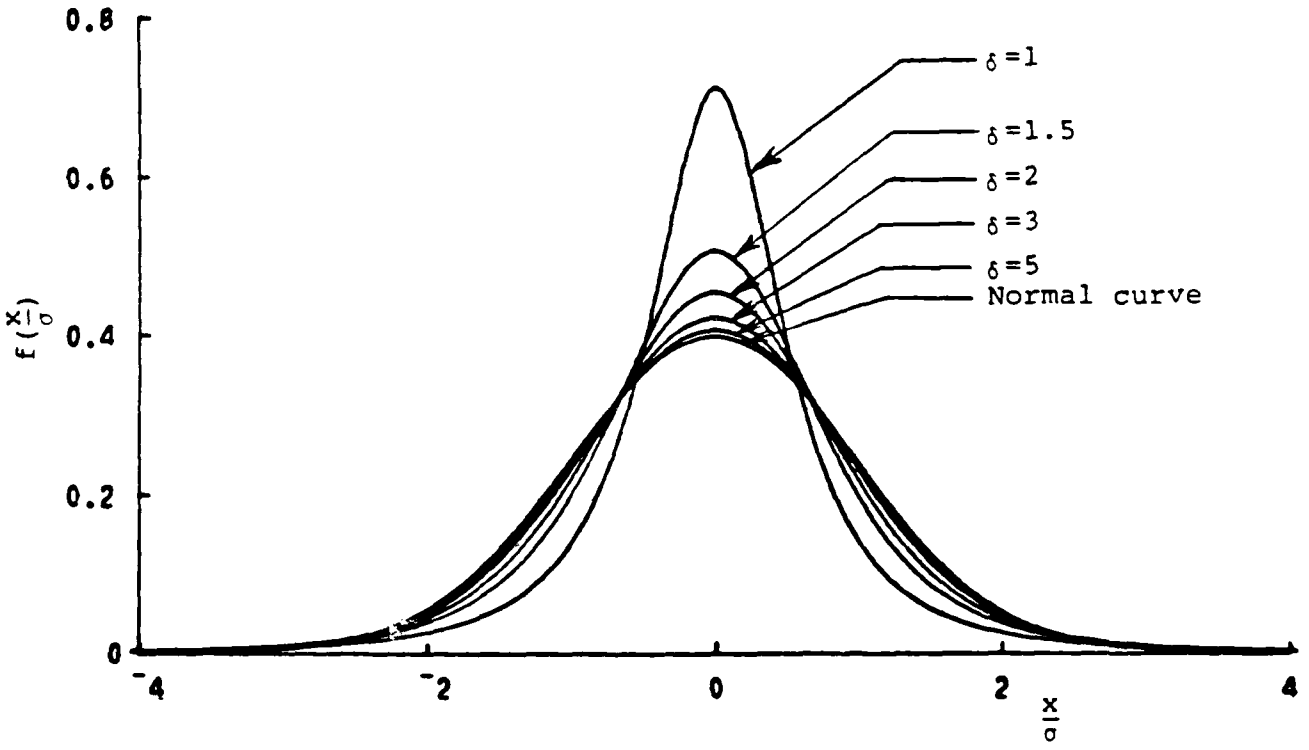


Fig. 2.13 - Johnson's S_u System: Density Functions

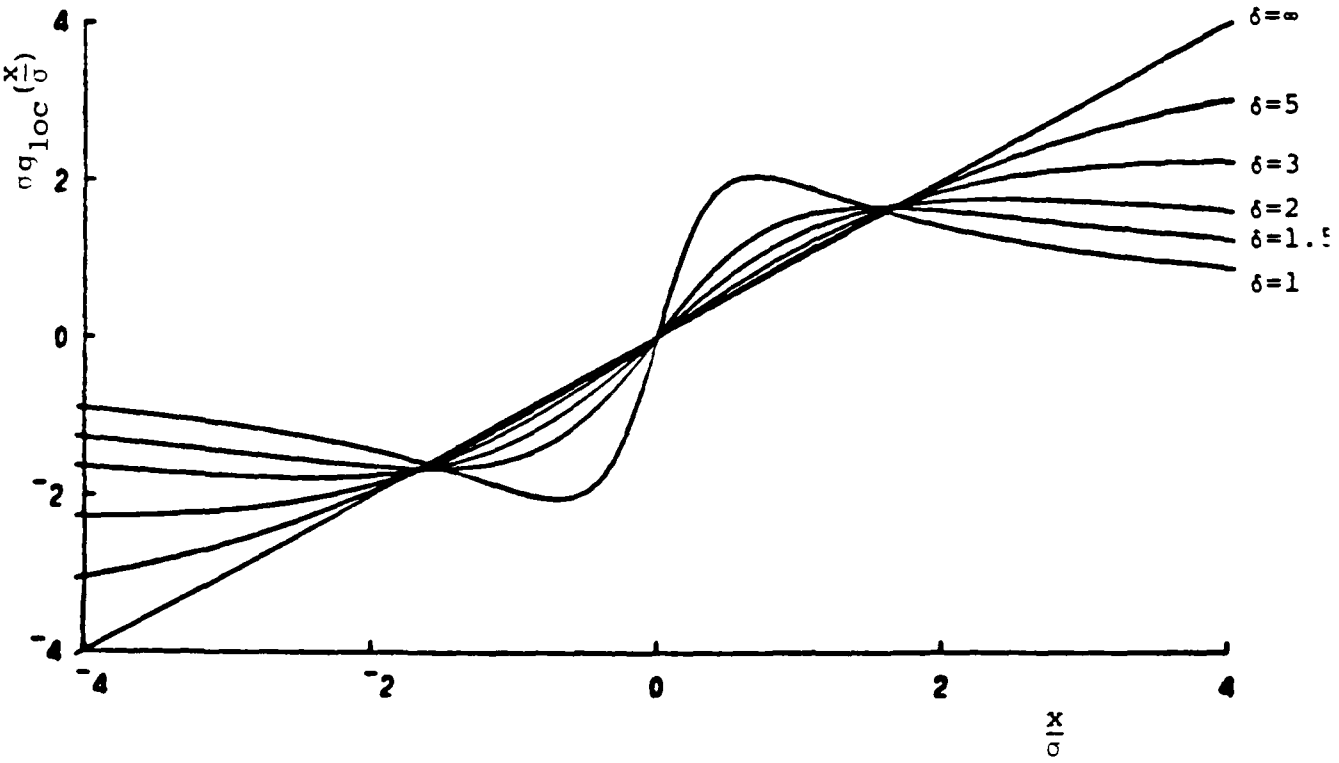


Fig. 2.14 - Johnson's S_u System: Locally Optimum Nonlinearities

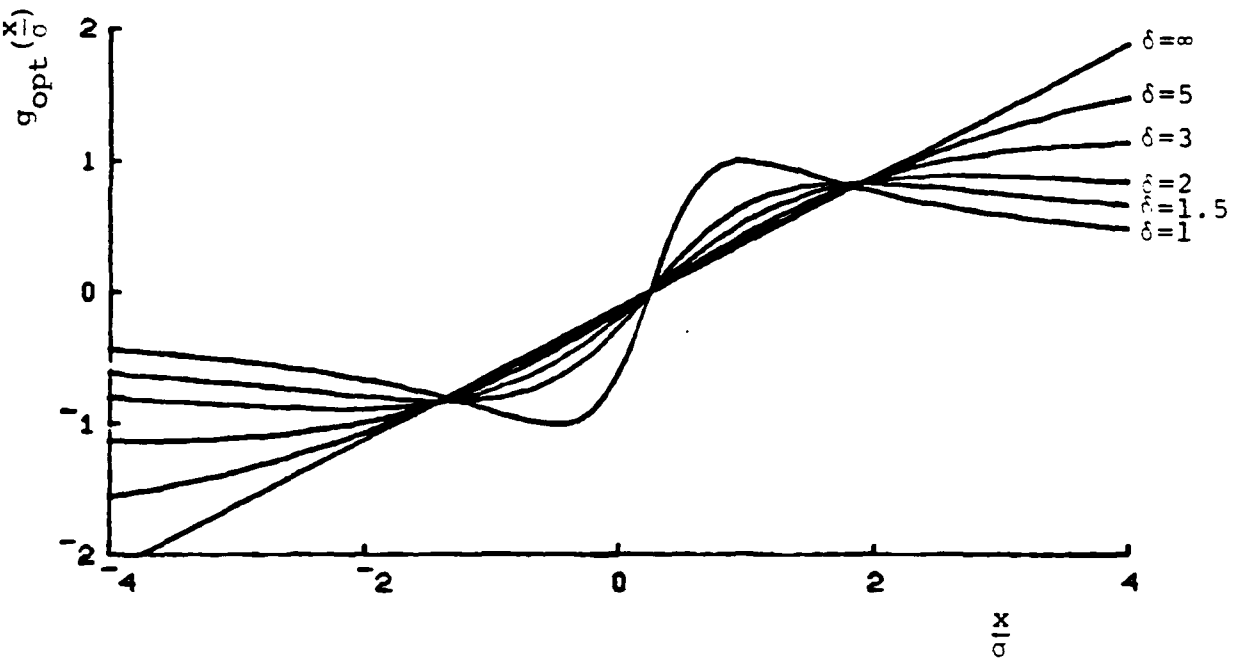


Fig. 2.15 - Johnson's S_u System: Optimum Nonlinearities (signal strength = $\frac{\sigma}{2}$)

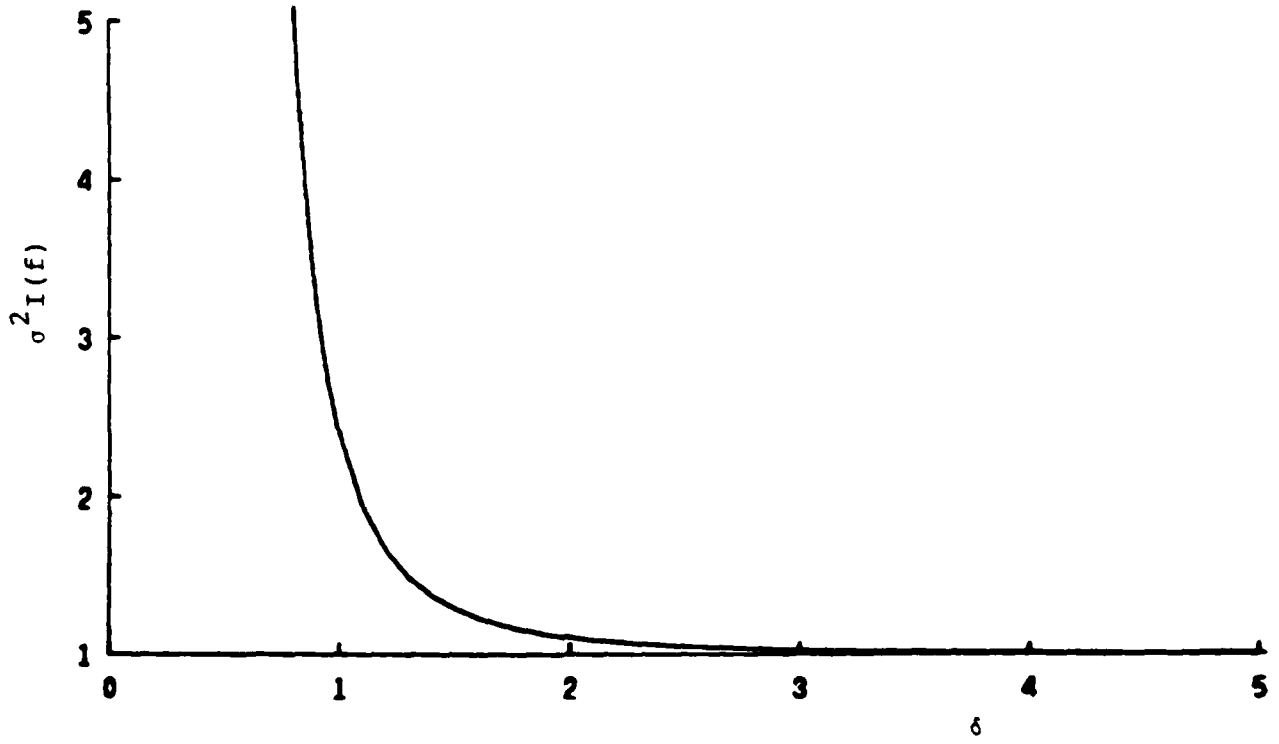


Fig. 2.16 - Johnson's S_u System: Fisher's Information vs. δ

The nonlinearities corresponding to the densities of Fig. 2.13 are given in Fig. 2.14. All locally optimum nonlinearities in this system are noise blankers since $\lim_{x\to\infty} g_{loc}(x) = 0$ for all δ [see Appendix 2.2]. The Johnson S_u System is an especially useful noise model since its distributions appear Gaussian in the middle but have heavier tails. That is, the locally optimum nonlinearities are nearly linear about the origin, but all become noise blankers eventually. The Neyman-Pearson optimum nonlinearities for $\theta=\sigma/2$ are plotted in Fig. 2.15.

A graph of Fisher's Information versus the parameter δ is given in Fig. 2.16. Small values of δ, corresponding to heavy-tailed distributions with large peaks at the origin, have relatively high values for Fisher's Information. As δ increases $\sigma^2 I(f)$ rapidly approaches one, the value for the Gaussian distribution.

Mixture Models

Mixtures have been frequently used in data analysis to either add uncertainty to statistical assumptions or to account for gross errors [24,25,26,27]. They have also been used in detection problems to test the robustness of standard detectors and to aid in the design of more robust detectors [28,29,30]. The basic mixture density is

$$f(x) = (1 - \epsilon)f_1(x) + \epsilon f_2(x) \qquad 0 < \epsilon < 1 \tag{2.38}$$

The density $f_1(x)$ is the nominal density and is very often chosen to be the Gaussian density. The density $f_2(x)$ is the contaminant. It is selected to fit the particular application. Huber [26] allows $f_2(x)$ to be a member of the class of all distributions in his work on robust estimation. Others [24,25] have selected $f_2(x)$ to be Gaussian with a variance much larger (variance ratios between 2 and 100 have been used) than that of $f_1(x)$. The percent of contamination ϵ is usually chosen to be small (often less than 0.1).

Figs. 2.17-2.24 display the densities, detector nonlinearities, and Fisher's Information for a mixture of two Gaussian distributions. For $f_1(x)$, the Gaussian density with variance $\sigma_1^2 = 1/(1-\epsilon+\epsilon\gamma^2)$ is used. The Gaussian density with variance $\sigma_2^2=\gamma^2\sigma_1^2$ is used for $f_2(x)$. Thus, the ratio is $\sigma_2^2/\sigma_1^2 = \gamma^2$ and the overall variance is unity since $\sigma^2=(1-\epsilon)\sigma_1^2+\epsilon\sigma_2^2=1$. Figs. 2.17, 2.19, and 2.21 display the densities, the locally optimum nonlinearities and the optimum nonlinearities for the Gaussian-Gaussian mixture with $\gamma^2=100$ and various values for ϵ. Figs. 2.18, 2.20, and 2.22 are similar graphs for $\epsilon=0.1$ and various γ^2 values. Note that the high-probability, low-amplitude portion of the nonlinearities is linear. This region is primarily determined by the nominal density, which is Gaussian. The low-probability, high-amplitude portion is also linear, but with a different slope. This region is primarily determined by the contaminating density. There is also a nonlinear transition region which connects these lines. Figs. 2.23 and 2.24 display Fisher's Information versus ϵ for $\gamma^2 = 100$ and Fisher's

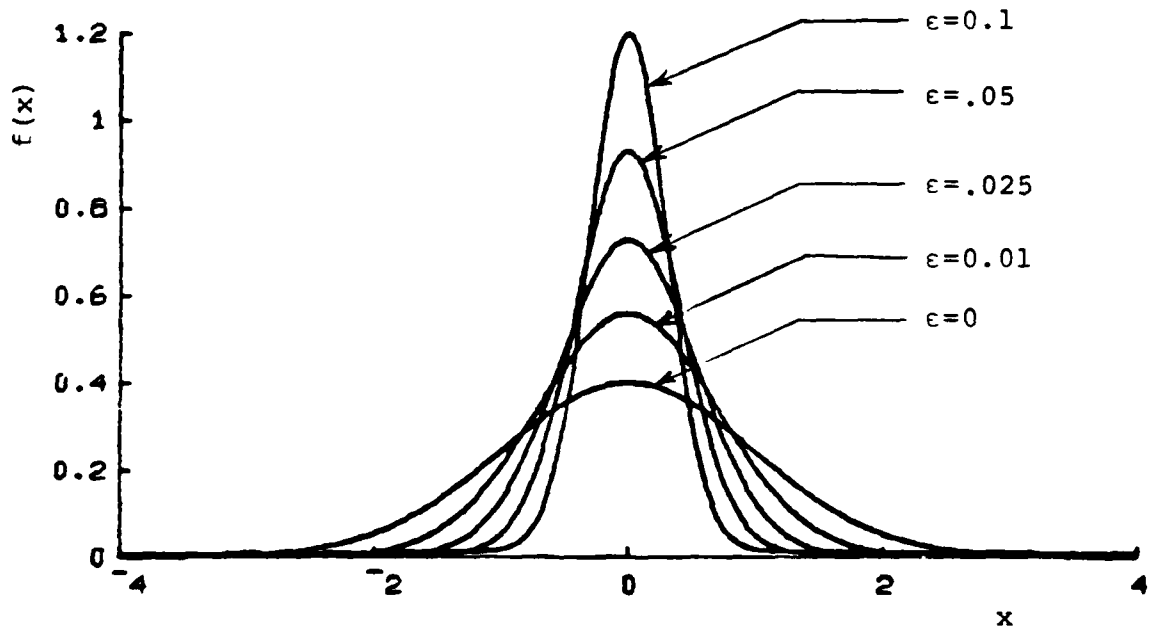


Fig. 2.17 - Gaussian-Gaussian Mixture: Density Functions
$\gamma^2=100$, $\sigma^2=1$

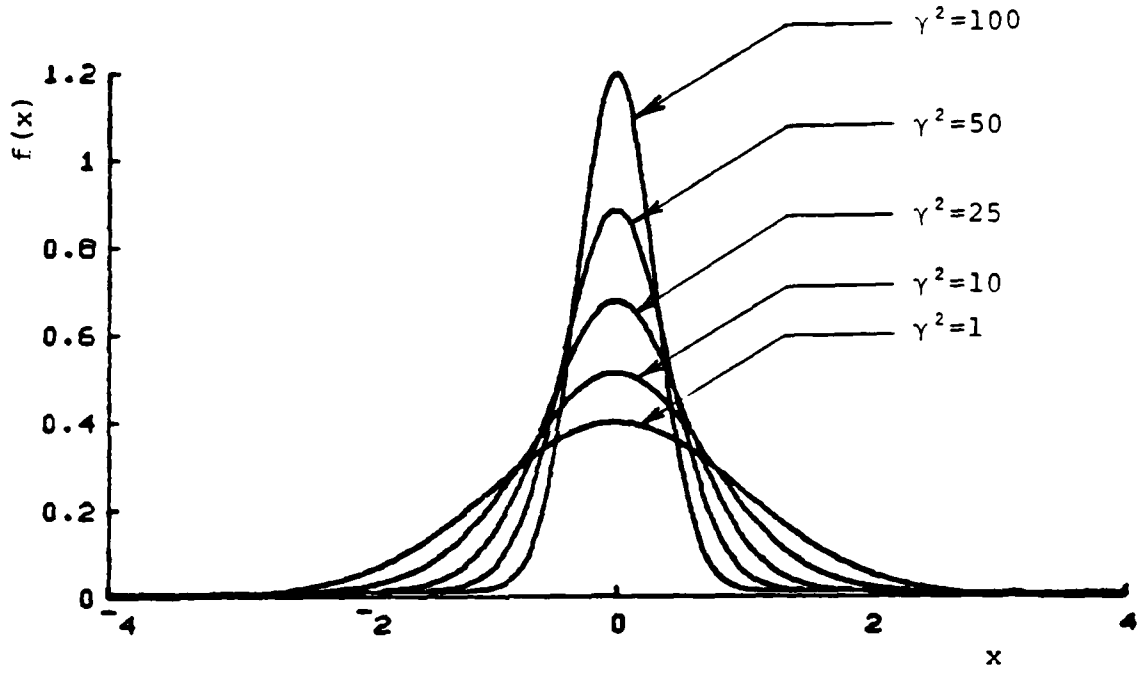


Fig. 2.18 - Gaussian-Gaussian Mixture: Density Functions
ε=0.1, σ^2=1

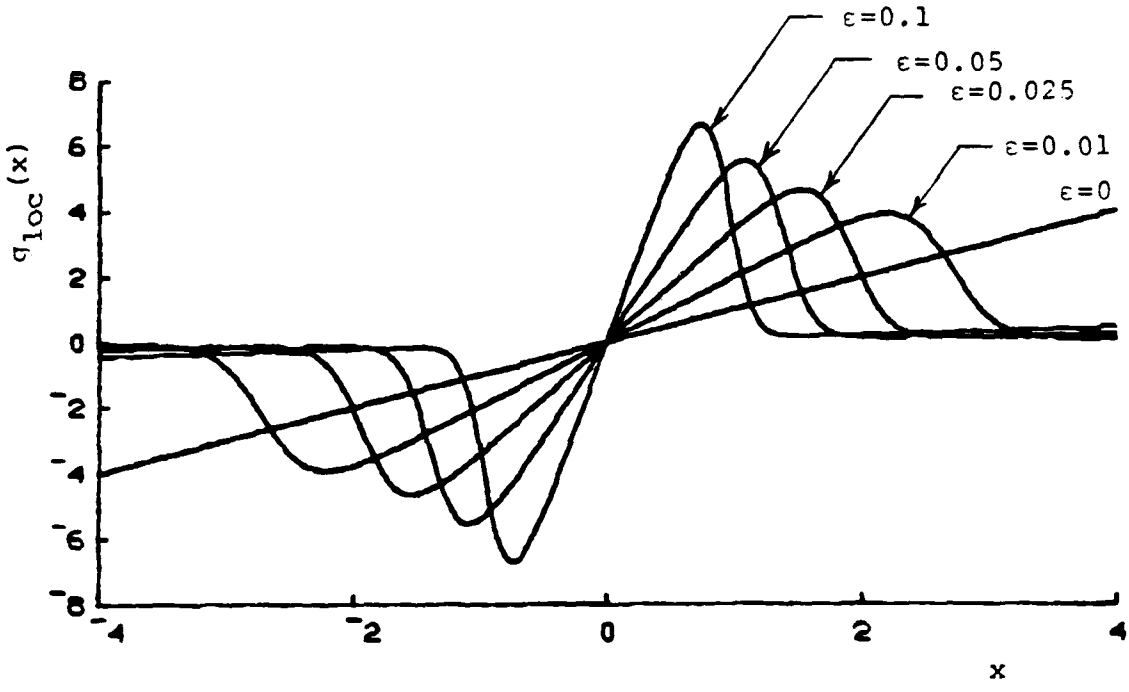


Fig. 2.19 - Gaussian-Gaussian Mixture: Locally Optimum Nonlinearities
$\gamma^2=100, \sigma^2=1$

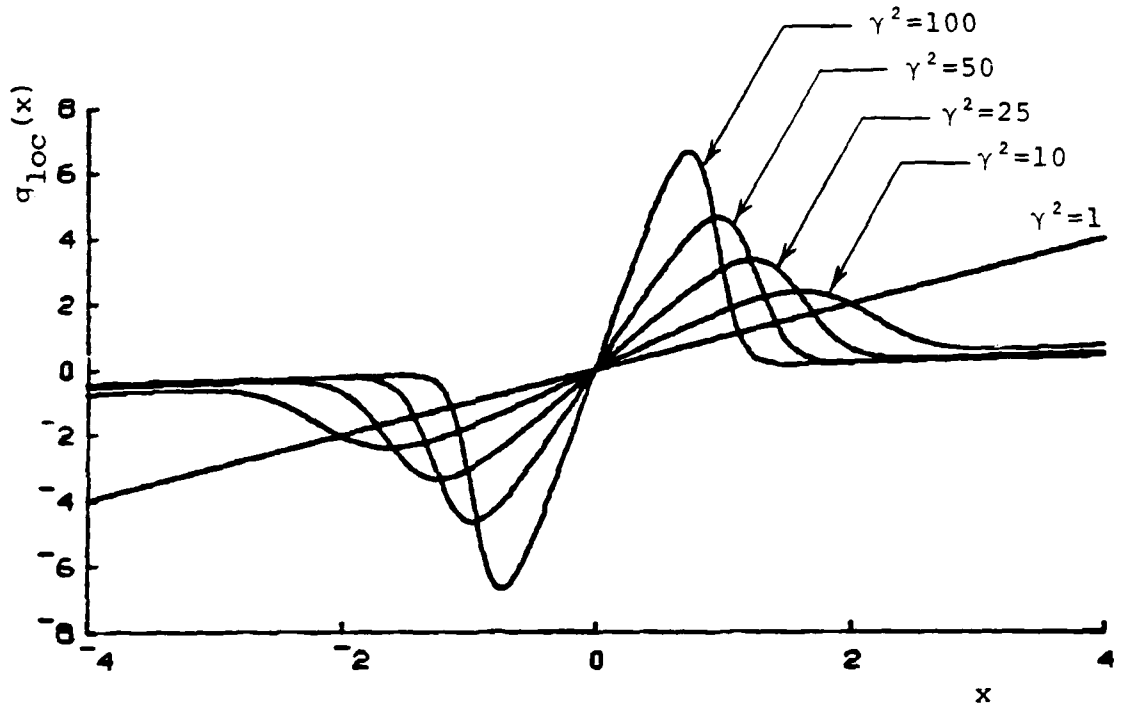


Fig. 2.20 - Gaussian-Gaussian Mixture: Locally Optimum Nonlinearities
ε=0.1, σ^2=1

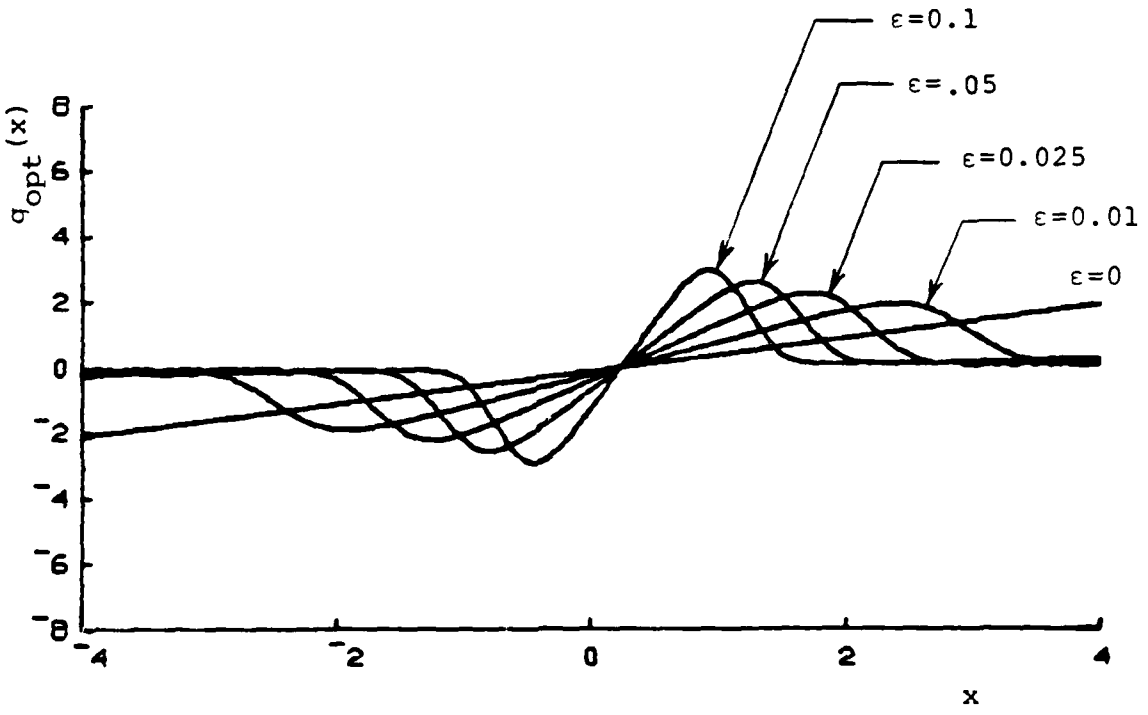


Fig. 2.21 - Gaussian-Gaussian Mixture: Optimum Nonlinearities
$\gamma^2=100$, $\sigma^2=1$, signal strength = 0.5

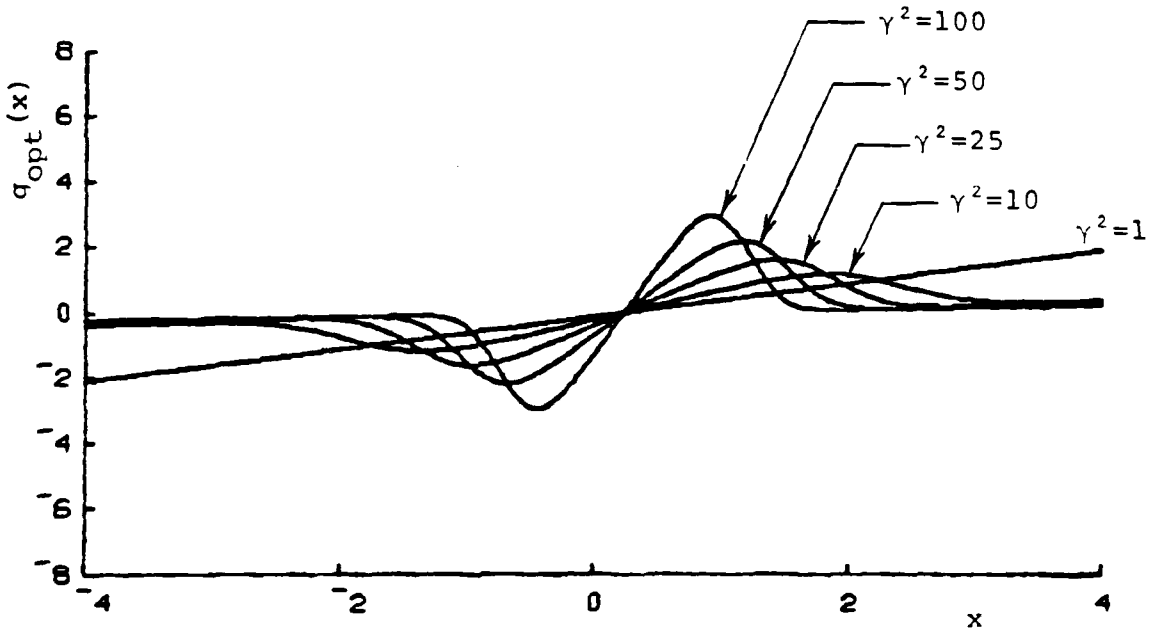


Fig. 2.22 - Gaussian-Gaussian Mixture: Optimum Nonlinearities
ϵ=0.1, σ^2=1, signal strength = 0.5

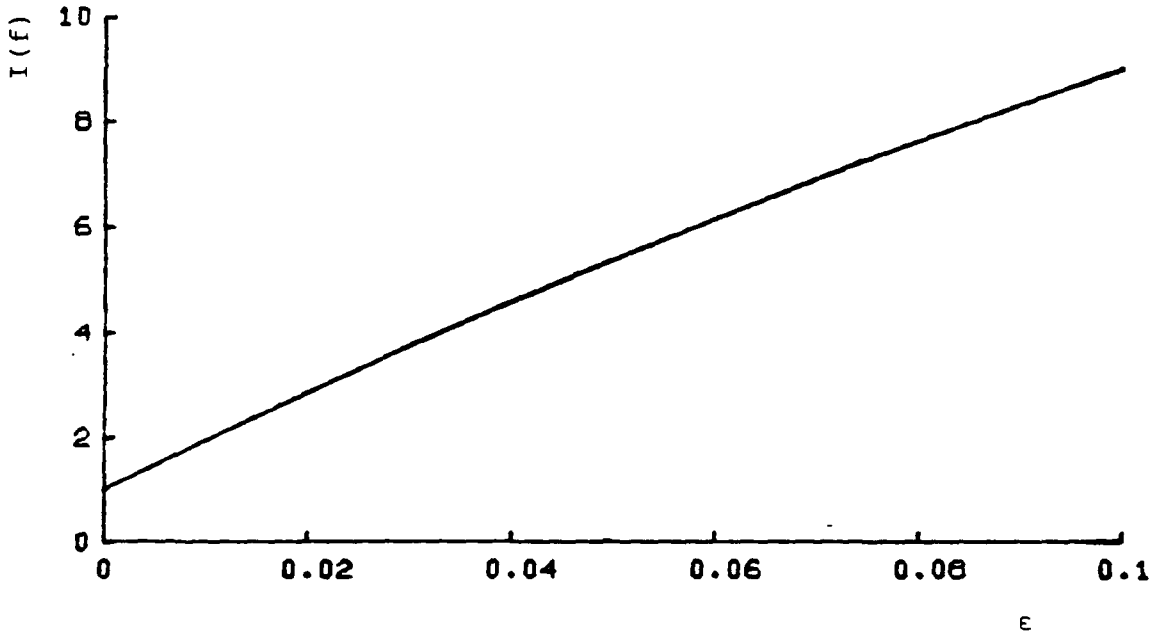


Fig. 2.23 - Gaussian-Gaussian Mixture: Fisher's Information vs. ε
$\gamma^2=100, \sigma^2=1$

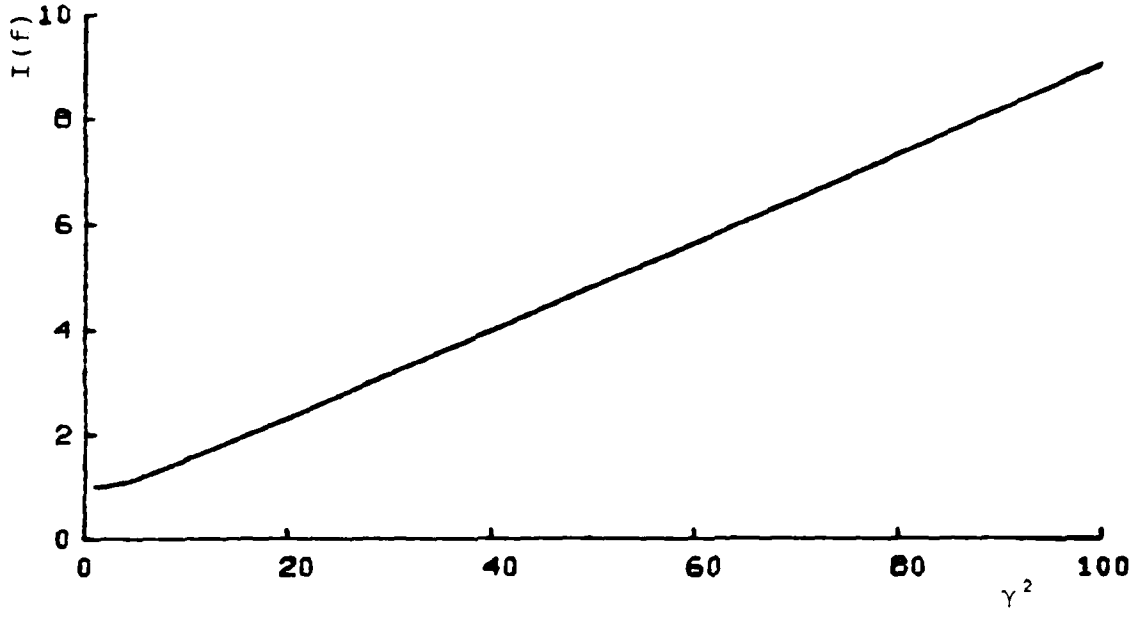


Fig. 2.24 - Gaussian-Gaussian Mixture: Fisher's Information vs. γ^2
$\varepsilon=0.1$, $\sigma^2=1$

Information versus γ^2 for $\varepsilon = 0.1$. These are not straight lines. However, over the range of most interest, the relationships are nearly linear.

Section 4: The Detection Problem in an Impulsive Environment

The detection problem introduced in Chapter I [Eq. (1.1)] can be solved if the statistical characteristics of the noise are known. The optimum detector is given by Eq. (1.5) and the locally optimum detector is given by Eq. (1.9). Both of these detectors consist of a nonlinearity followed by a summer and a threshold comparator, as shown in Fig. 1.1. The difficulty of course is that the density function for the noise is not perfectly known. Equations have been formulated (see Chapter II - Sections 2 and 3) to model noise; however, it must be remembered that these are only approximations to the true densities. The optimum detectors for these models will not necessarily be optimum for the actual noise. Worse yet, the noise is often non-stationary. The optimum detector would then have to change to reflect the changes in the noise statistics.

Since optimum nonlinearities are difficult, if not impossible, to design for actual noise, several authors [29,31,32,33] have studied the performance of different suboptimum nonlinearities. Bernstein [31] calculated the locally optimum nonlinearity for a set of VLF atmospheric noise data. He reported that the nonlinearity was roughly linear near the origin and that it suppressed high amplitude

observations. Both limiters and blankers can be used as approximations to this type of nonlinearity. Fig. 2.25 displays simple limiter and blanker nonlinearities. The performance of limiters and blankers for various noise models has been studied [29,31,32,33]. The noise blanker (or hole puncher) performs well for a wide variety of densities; however, its performance is very sensitive to the selection of the noise blanking level. The limiter (or clipper), although not quite as effective as the blanker, is less sensitive to the selection of its limiting level. Other more complex approximations to the optimum nonlinearities may be used [29]. The better the approximation, the better will be the performance. However, the disadvantages of using a complex nonlinearity in the receiver may outweigh the increased performance.

Another approach to the problem of detecting signals when the statistics of the noise are not well known is through the techniques of robust detection. Often the noise can be considered to belong to a well defined class of noises. One then must look for a detector which has reasonably good performance over the entire class. Several authors [30,34,35,36] have considered different classes of mixture models. In many cases mini-max detectors have been established for the class. Some results which facilitate the design of mini-max detectors are presented in Chapter III of this dissertation.

If very little information is available on the statistical characteristics of the noise, it is often wise to use

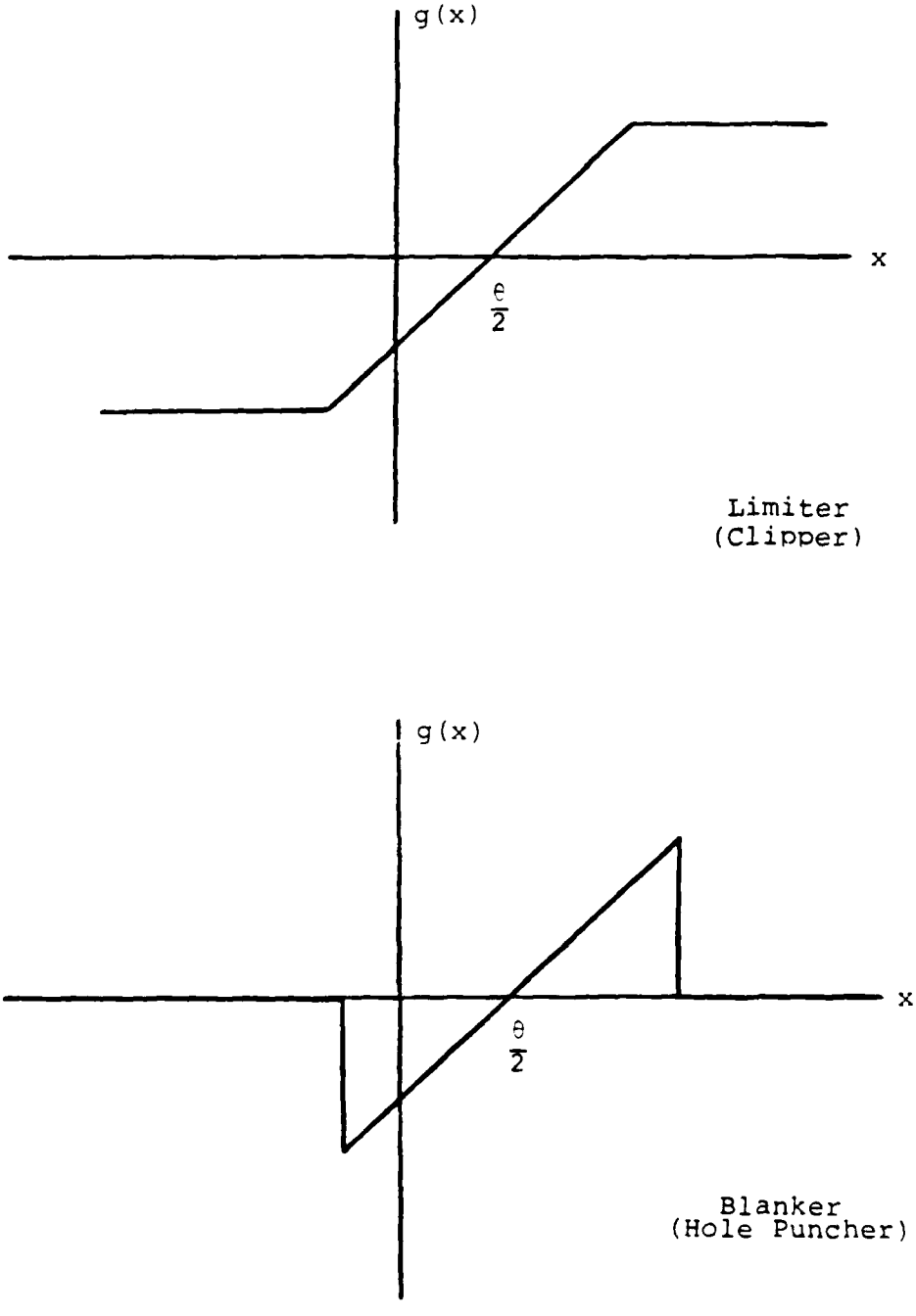


Fig. 2.25 - Limiter and Blanker Nonlinearities

nonparametric detection techniques. There are many good references on nonparametric detection [37,38,39,40]. These techniques will not be treated further in this dissertation.

For applications where the noise statistics are unknown or time-varying and near optimum detector performance is essential, an adaptive detector may be necessary. Although more complex than those detectors mentioned above, when properly designed, it can be nearly optimum for a wide variety of noise environments [41,42,43]. Chapter IV presents an adaptive detection scheme which performs well for several different noise densities.

Appendix 2.1

Fisher's Information for the Generalized Gaussian Distribution of Chapter II - Section 3

Starting with Eq. (1.17) and substituting Eqs. (2.30) and (2.31), we have

$$I(f) = \int_{-\infty}^{\infty} \left(\frac{f'(x)}{f(x)} \right)^2 f(x)\,dx$$

$$= \int_{-\infty}^{\infty} (c\eta^c\,|x|^{c-1}\,\mathrm{sgn}(x)\,)^2\,\frac{c\eta}{2\Gamma(\frac{1}{c})}\,e^{-[\eta|x|]^c}\,dx, \quad c > 0$$

where $\eta = \frac{1}{\sigma}\left[\frac{\Gamma(\frac{3}{c})}{\Gamma(\frac{1}{c})}\right]^{1/2}$

This can be simplified to

$$I(f) = \frac{c^3\eta^{2c+1}}{\Gamma(\frac{1}{c})} \int_0^{\infty} x^{2c-2}\,e^{-\eta^c x^c}\,dx, \qquad c > 0$$

Let $t = \eta^c x^c$ and obtain

$$I(f) = \frac{c^2\eta^2}{\Gamma(\frac{1}{c})} \int_0^{\infty} t^{(2-\frac{1}{c})-1}\,e^{-t}\,dt, \qquad c > 0$$

For $c > 1/2$, the integral equals $\Gamma(2-\frac{1}{c})$ and thus

$$I(f) = \frac{c^2\eta^2}{\Gamma(\frac{1}{c})}\,\Gamma(2-\tfrac{1}{c}) = \frac{c^2}{\sigma^2}\,\frac{\Gamma(\frac{3}{c})\,\Gamma(2-\frac{1}{c})}{\left[\Gamma(\frac{1}{c})\right]^2}, \qquad c > 1/2$$

For $0 < c \leq 1/2$, let $\alpha = 2-\frac{1}{c}$. Then $0 < c \leq 1/2 \Rightarrow \alpha \leq 0$ and

$$I(f) = \frac{\eta^2}{(2-\alpha)^2\Gamma(2-\alpha)} \int_0^\infty t^{\alpha-1} e^{-t}\, dt$$

However,

$$\int_0^\infty t^{\alpha-1} e^{-t}\, dt \geq \int_0^1 t^{\alpha-1} e^{-t}\, dt$$
$$\geq e^{-1}\int_0^1 t^{\alpha-1}\, dt$$

and $\int_0^1 t^{\alpha-1}\, dt = \infty$ for $\alpha \leq 0$.

Thus $I(f) = \infty$ for $0 < c \leq 1/2$. This establishes Eq.(2.33).

Appendix 2.2

A Proof that $\lim_{x\to\infty} g_{loc}(x) = 0$ for Johnson's S_u System

One can write

$$g_{loc}(x) = \left[1 + \left(\frac{x}{\lambda}\right)^2\right]^{-1} \left(\frac{x}{\lambda_2}\right) + \left[1 + \left(\frac{x}{\lambda}\right)^2\right]^{-1/2} \frac{\delta^2}{\lambda} \sinh^{-1}\left(\frac{x}{\lambda}\right)$$

$$= \frac{x}{\lambda^2+x^2} + \frac{\delta^2 \sinh^{-1}(x/\lambda)}{\lambda\left[1 + x^2/\lambda^2\right]^{1/2}}$$

Then,

$$\lim_{x\to\infty} g_{loc}(x) = \lim_{x\to\infty} \frac{\delta^2 \sinh^{-1}(x/\lambda)}{\lambda\left[1 + x^2/\lambda^2\right]^{1/2}}$$

$$= \lim_{x\to\infty} \frac{\delta^2}{x} = 0$$

by L'Hopital's rule.

References - Chapter II

1. A. D. Watt and E. L. Maxwell, "Characteristics of Atmospheric Noise from 1 to 100 KC," Proc. IRE, Vol. 45, pp. 787-794, June 1957.

2. A. D. Watt and E. L. Maxwell, "Measured Statistical Characteristics of VLF Atmospheric Radio Noise," Proc. IRE, Vol. 45, pp. 55-62, January 1957.

3. A. D. Watt, V. L. F. Radio Engineering, Pergamon Press, New York, 1967.

4. H. M. Hall, "A New Model for 'Impulsive' Phenomena: Application to Atmospheric-Noise Communication Channels," Tech. Rep. 3412-8, Stanford Elec. Lab., Stanford, CA, August 1966.

5. V. O. Knudsen, R. S. Alford, and J. W. Emling, "Underwater Ambient Noise," Journal of Marine Research, Vol. 7, pp. 410-429, 1948.

6. G. M. Wenz, "Acoustic Ambient Noise in the Ocean: Spectra and Sources," Journal Acoustical Soc. Amer., Vol. 34, No. 12, pp. 1936-1956, December 1962.

7. A. R. Milne and J. H. Ganton, "Ambient Noise under Arctic Sea Ice," Journal Acoustical Soc. Amer., Vol. 36, No. 5, pp. 855-863, May 1964.

8. A. D. Spaulding and D. Middleton, "Optimum Reception in an Impulsive Interference Environment," OT Report 75-67, Office of Telecommunications, U.S. Dept. of Commerce, June 1975.

9. E. N. Skomal, "Distribution and Frequency Dependence of Unintentionally Generated Man-Made VHF/UHF Noise in Metropolitan Areas. Part II - Theory," IEEE Trans. on Electromagnetic Compatibility, Vol. EMC-7, pp. 420-427, December 1965.

10. P. Mertz, "Model of Impulsive Noise for Data Transmission," IRE Trans. Comm. Systems, Vol. CS-9, No. 2, pp. 130-137, June 1961.

11. J. S. Engel, "Digital Transmission in the Presence of Impulsive Noise," Bell Sys. Tech. Journal, Vol. 44, No. 8, pp. 1699-1743, October 1965.

12. W. Q. Crichlow, C. J. Roubique, A. D. Spaulding, and W. M. Beery, "Determination of the Amplitude-Probability Distribution of Atmospheric Radio Noise from Statistical Moments," Journal of Research NBS, Vol. 64D (Radio Prop.), No. 1, pp. 49-56, January 1960.

13. F. Horner and J. Harwood, "An Investigation of Atmospheric Radio Noise at Very Low Frequencies," Proc. Inst. Elec. Engrs., No. 103, Pt. B, pp. 743-751, November 1956.

14. P. Beckmann, "Amplitude-Probability Distribution of Atmospheric Radio Noise," Journal of Research NBS, Vol. 68D (Radio Science), No. 6, pp. 723-736, June 1964.

15. A. D. Spaulding, C. J. Roubique, and W. Q. Crichlow, "Conversion of the Amplitude-Probability Distribution Function for Atmospheric Radio Noise from One Bandwidth to Another," Journal of Research NBS, Vol. 66D (Radio Prop.), No. 6, pp. 713-720, November 1962.

16. K. Furutsu and T. Ishida, "On the Theory of Amplitude Distribution of Impulsive Random Noise," J. of Applied Physics, Vol. 32, No. 7, pp. 1206-1221, July 1961.

17. A. A. Giordano and F. Haber, "Modeling of Atmospheric Noise," Radio Science, Vol. 7, No. 11, pp. 1011-1023, November 1972.

18. D. Middleton, "Man-Made Noise in Urban Environments and Transportation Systems: Models and Measurements," IEEE Trans. Comm., Vol. COM-21, No. 11, pp. 1232-1241, November 1973.

19. A. D. Spaulding and D. Middleton, "Optimum Reception in an Impulsive Interference Environment - Part I: Coherent Detection," IEEE Trans. Comm., Vol. COM-25, No. 9, pp. 910-923, September 1977.

20. A. D. Spaulding and D. Middleton, "Optimum Reception in an Impulsive Interference Environment - Part II: Incoherent Reception," IEEE Trans. Comm., Vol. COM-25, No. 9, pp. 924-934, September 1977.

21. M. Abramowitz and I. A. Stegun ed., Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, National Bureau of Standards Applied Mathematics Series #55, pg. 504, Eq. 13.1.2, Dec. 1972.

22. M. Kanefsky and J. B. Thomas, "On Polarity Detection Schemes with Non-Gaussian Inputs," J. Franklin Institute, Vol. 280, pp. 120-138, August 1965.

23. N. L. Johnson, "Systems of Frequency Curves Generated by Methods of Translation," Biometrika, Vol. 36, pp. 149-176, 1949.

24. J. W. Tukey, "A Survey of Sampling from Contaminated Distributions," in Contributions to Probability and Statistics, I. Olkin, ed., Stanford Univ. Press, pp. 448-485, 1960.

25. D. F. Andrews, P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, J. W. Tukey, Robust Estimates of Location: Survey and Advances, Princeton University Press, Princeton, NJ, 1972.

26. P. J. Huber, "Robust Estimation of a Location Parameter," Ann. Math. Stat., Vol. 35, pp. 73-101, March 1964.

27. P. J. Huber, "Robust Statistics: A Review," Ann. Math. Stat., Vol. 43, No.4, pp. 1041-1067, August 1972.

28. J. H. Miller and J. B. Thomas, "The Detection of Signals in Impulsive Noise Modeled as a Mixture Process," IEEE Trans. Communications, Vol. COM-24, pp. 559-563, May 1976.

29. J. H. Miller and J. B. Thomas, "Robust Detectors for Signals in Non-Gaussian Noise," IEEE Trans. Communications, Vol. COM-25, No. 7, pp. 686-690, July 1977.

30. A. H. El-Sawy and V. D. VandeLinde, "Detection of Signals in the Presence of Impulsive Noise," Proc. 1978 Conf. on Infor. Sciences and Systems, Johns Hopkins University, pp. 7-12, 1978.

31. S. L. Bernstein, et. al., "Long-Range Communications at Extremely Low Frequencies," Proc. IEEE, Vol. 62, No. 3, pp. 292-312, March 1974.

32. R. F. Ingram and R. Houle, "Performance of the Optimum and Several Suboptimum Receivers for Threshold Detection of Known Signals in Additive, White, Non-Gaussian Noise," NUSC Technical Report 6339, Naval Underwater Systems Center, New London, CT, November 1980.

33. R. E. Ziemer and R. B. Fluchel, "Selection of Blanking and Limiting Levels for Binary Signaling in Gaussian Plus Impulse Noise," Proc. of the IEEE 1971 Fall Electronics Conference, Chicago, Illinois, pp. 290-295, October 1971.

34. R. D. Martin and S. C. Schwartz, "Robust Detection of a Known Signal in Nearly Gaussian Noise," IEEE Trans. Infor. Theo., Vol. IT-17, No. 1, January 1971.

35. S. A. Kassam and J. B. Thomas, "Asymptotically Robust Detection of a Known Signal in Contaminated Non-Gaussian Noise," IEEE Trans. Infor. Theo., Vol. IT-22, No. 1, pp. 22-26, January 1976.

36. A. H. El-Sawy and V. D. VandeLinde, "Robust Detection of Known Signals," IEEE Trans. Infor. Theo., Vol. IT-23, No. 6, pp. 722-727, November 1977.

37. S. A. Kassam and J. B. Thomas, ed., Nonparametric Detection Theory and Applications, Dowden, Hutchinson and Ross, Inc., Stroudsburg, PA, 1980.

38. D. A. S. Fraser, Nonparametric Methods in Statistics, John Wiley and Sons, Inc., New York, 1957.

39. P. Papantoni-Kazakos and D. Kazakos, ed., Nonparametric Methods in Communications, Marcel Dekker Inc., New York, 1977.

40. J. Gibson and J. Melsa, Introduction to Nonparametric Detection with Applications, Academic Press, New York, 1975.

41. H. L. Groginsky, L. R. Wilson, and D. Middleton, "Adaptive Detection of Statistical Signals in Noise," IEEE Trans. Infor. Theo., Vol. IT-12, No. 3, pp. 337-348, July 1966.

42. L. D. Davisson, " A Theory of Adaptive Filtering," IEEE Trans. Infor. Theo., Vol IT-12, No. 2, pp. 97-102, April 1966.

43. J. W. Modestino, "Adaptive Detection of Signals in Impulsive Noise Environments," IEEE Trans. on Comm., Vol. COM-25, No. 9, pp. 1022-1027, September 1977.

Chapter III

MINIMAX DETECTORS

In many detection problems the statistical characterization of the noise is not complete. For example, it may only be possible to define a class of probability densities which contains the actual noise density. In such cases minimax detectors may be used to guarantee a lower bound on detector performance for the entire class.

Once a class of densities has been specified, the worst case density within the class must be identified. The best detector for this density is the minimax detector for the class. Sections 1, 2, and 3 of this chapter establish the pair (f_o, D_o) as the saddlepoint of this detection problem. The density f_o is the member of the class with minimum Fisher's Information. The detector D_o is the locally optimum detector for density f_o. In Section 4 a few classes of densities are considered and relationships are established to determine the minimax detector.

Section 1: The Relationship Between Fisher's Information and the Asymptotic Performance of Optimum Detectors

Fisher's Information has been associated with the asymptotic performance of estimators and detectors before. Capon [1] has related it to the efficacy of the locally optimum detector (see Appendix 3.1). Ingram [2] related Fisher's Information to the improvement (measured as an increase in the signal-to-noise ratio) obtained with using the locally optimum detector rather than the linear detector. Huber [3] related Fisher's Information to the asymptotic variance of estimators in his work on robust estimation. He concluded that the most robust M-estimate over a given class of densities is the maximum likelihood estimate for the density which minimizes Fisher's Information. El-Sawy and VandeLinde [4] applied Huber's results to detection problems. Their M-detectors utilize Huber's M-estimates to form the test statistic.

The results presented here are different. Fisher's Information is related directly to the asymptotic probability of detection for both optimum and locally optimum detectors. It forms part of a generalized signal-to-noise power ratio and can be thought of as a measure of the difficulty of detecting a signal in a given noise.

Consider the problem which was first presented in Eq. (1.1),

$$\begin{aligned} H:&\quad x_i = n_i \\ K:&\quad x_i = n_i + \theta s_i \end{aligned} \qquad \theta > 0,\ i=1,\ldots,N \tag{3.1}$$

As before, the known signals $\{\theta s_i\}_1^N$ obey $0 < m_1 < |s_i| < M_1 < \infty$. The noise samples $\{n_i\}_1^N$ are realizations of the independent random variables $\{N_i\}_1^N$ with densites $f_i \in F$, a class of symmetric densites which will be defined below. The Neyman-Pearson optimum detector is given by Eq. (1.5) and the locally optimum detector is given by Eq. (1.9).

The performance of these detectors is usually measured by the power function $\beta(\theta) = \text{Prob}_K(\text{decide K})$, and by the level $\alpha = \text{Prob}_H(\text{decide K})$. For a fixed level α, the Neyman-Pearson optimum detector maximizes the power function $\beta(\theta)$, while the locally optimum detector maximizes the slope of the power function near the origin $\left(\frac{d\beta}{d\theta}\Big|_{\theta\to 0}\right)$. The level α is fixed by the appropriate choice of the threshold T. In order to calculate the power function, the probability distribution function of the test statistic [$T_{opt}(\underline{x})$ of Eq. (1.5) or $T_{loc}(\underline{x})$ of Eq. (1.9)] is required under both H and K. This is difficult to obtain in general since an N-fold convolution of the densities $\{f_i\}_1^N$ is required. However, a few special cases have been solved. The performance of the linear detector in Gaussian noise is well known and the amplifier-limiter operating in Laplace noise has also been studied recently [5].

In detecting a small signal using a large number of observations, it is usually possible to use the Central Limit Theorem to show that the test statistic is asympotically normally distributed. One simple form of the Central Limit Theorem [6] states

Let $t_1,\ldots,t_N$ be a series of independent random variables having arbitrary distributions, means $\mu_1,\ldots,\mu_N$ and variances $\sigma^2{}_1,\ldots,\sigma^2{}_N$ respectively. Let

$$y_n = \frac{\sum_{i=1}^{N} t_i - \sum_{i=1}^{N} \mu_i}{\left(\sum_{i=1}^{N} \sigma_i^2\right)^{1/2}} \qquad , n = 1,\ldots,N$$

If $\sigma^2{}_i < \infty$ for all i and if there exists positive constants m and M such that $\sigma^2{}_i > m$ and $E\{|t_i-\mu_i|^3\} < M$ for all i, then the distribution of y_n approaches in the limit the unit normal distribution. (In the special case where the $\{t_i\}_1^N$ are identically distributed, the condition $\sigma^2{}_i < \infty$ for all i is sufficient.)

Using this theorem, it is straightforward to show that

$$\beta(\theta) \to 1 - \Phi\left(\frac{T - \sum_{i=1}^{N} \mu_{iK}}{\left(\sum_{i=1}^{N} \sigma_{iK}^2\right)^{1/2}}\right) \tag{3.2}$$

and

$$\alpha \to 1 - \Phi\left(\frac{T - \sum_{i=1}^{N} \mu_{iH}}{\left(\sum_{i=1}^{N} \sigma_{iH}^2\right)^{1/2}}\right) \tag{3.3}$$

provided that the conditions on μ_i and $\sigma^2{}_i$ are met under both H and K. These three conditions are used to define the class of densities F for which the results of this section hold. The symbol $\to$ signifies equality in the limit as N approaches infinity and θ approaches zero. The symbol $\Phi(\cdot)$ represents the unit normal cumulative distribution function.

Also, $\mu_{iK} = E_K[g_i(x_i)]$ and $\sigma^2_{iK} = Var_K[g_i(x_i)]$. Similarly, μ_{iH} and σ^2_{iH} are the corresponding means and variances under H. For the Neyman-Pearson optimum detector $g_i(x_i) = g_{opt_i}(x_i)$ and for the locally optimum detector $g_i(x_i) = g_{loc_i}(x_i)$ [as defined by Eqs. (1.6) and (1.10)].

First consider the Neyman-Pearson optimum detector

$$T_{opt}(\underline{x}) = \sum_{i=1}^{N} \log \frac{f_i(x_i-\theta s_i)}{f_i(x_i)} \quad \begin{cases} > T \text{ decide K} \\ < T \text{ decide H} \end{cases} \tag{3.4}$$

Assume that the f_i have infinite support with derivatives of all order on $(-\infty,\infty)$ and that $f_i(\infty) = f_i(-\infty) = f_i'(\infty) = f_i'(-\infty) = 0$ for all i. In order to calculate the power function using the Gaussian approximation provided by the Central Limit Theorem, the mean and variance of the single sample test statistic must be calculated.

$$\mu_{iH} = E_H\left[g_{opt_i}(x_i) \right] = E_H\left[\log \frac{f_i(x_i-\theta s_i)}{f_i(x_i)} \right] \tag{3.5}$$

Expanding in a Taylor series about $\theta=0$ yields

$$\mu_{iH} = -\frac{1}{2}\theta^2 s_i^2 I(f_i) + \theta^4 K_1(\theta, s_i, f_i) \tag{3.6}$$

where

$$I(f_i) = \int_{-\infty}^{\infty} \left(\frac{f_i'(x)}{f_i(x)} \right)^2 f_i(x)\, dx \tag{3.7}$$

is Fisher's measure of information for the location parameter in the density $f_i(\cdot)$, and $K_1(\theta, s_i, f_i)$ is a power series containing the remaining terms of the expansion. Similarly,

$$\sigma_{iH}^2 = E_H\left[\log \frac{f_i(x_i-\theta s_i)}{f_i(x_i)}\right]^2 - \mu_{iH}^2$$

$$= \theta^2 s_i^2 I(f_i) + \theta^4 K_2(\theta, s_i, f_i) \tag{3.8}$$

where $K_2(\theta, s_i, f_i)$ contains the remaining terms of the series. Since $f_i(\cdot)$ is assumed symmetric, it can be shown [7] that $\mu_{iK} = -\mu_{iH}$ and $\sigma^2_{iK} = \sigma^2_{iH}$. Using the Gaussian approximation, the threshold is chosen to fix α:

$$\alpha = \text{Prob}_H\left[T_{opt}(\underline{x}) > T\right] \approx 1 - \Phi\left(\frac{T - \sum_{i=1}^{N} \mu_{iH}}{\left(\sum_{i=1}^{N} \sigma_{iH}^2\right)^{1/2}}\right) \tag{3.9}$$

$$T \approx \Phi^{-1}(1-\alpha)\left(\sum_{i=1}^{N} \sigma_{iH}^2\right)^{1/2} + \sum_{i=1}^{N} \mu_{iH} \tag{3.10}$$

The symbol $\approx$ denotes approximate equality. The power function is then calculated:

$$\beta(\theta) = \text{Prob}_K\left[T_{opt}(\underline{x}) > T\right] \approx 1 - \Phi\left(\frac{T - \sum_{i=1}^{N} \mu_{iK}}{\left(\sum_{i=1}^{N} \sigma_{iK}^2\right)^{1/2}}\right) \tag{3.11}$$

$$\beta(\theta) \approx 1 - \Phi\left(\Phi^{-1}(1-\alpha) + \frac{2\sum_{i=1}^{N} \mu_{iH}}{\left(\sum_{i=1}^{N} \sigma_{iH}^2\right)^{1/2}}\right) \tag{3.12}$$

In the limit as $\theta \to 0$ and $N \to \infty$ with $k^2 = \theta^2 \sum_{i=1}^{N} s_i^2 I(f_i) =$ a nonzero constant, the power function becomes

$$\beta(\theta) = 1 - \Phi(\Phi^{-1}(1-\alpha) - k) \tag{3.13}$$

provided μ_{iH} and σ^2_{iH} are bounded.

The performance for the locally optimum detector is calculated similarly.

$$T_{loc}(\underline{x}) = \sum_{i=1}^{N} \frac{-f_i'(x_i)}{f_i(x_i)} s_i \quad \begin{cases} > T & \text{decide K} \\ < T & \text{decide H} \end{cases} \tag{3.14}$$

The mean and variance of the single sample test statistic under H are

$$\mu_{iH} = E_H\left(\frac{-f_i'(x_i)}{f_i(x_i)} s_i\right) = 0 \tag{3.15}$$

$$\sigma^2_{iH} = E_H\left(\frac{-f_i'(x_i)}{f_i(x_i)} s_i\right)^2 = s_i^2 I(f_i) \tag{3.16}$$

The mean under K is

$$\mu_{iK} = E_K\left(\frac{-f_i'(x_i)}{f_i(x_i)} s_i\right) = (-s_i)\int_{-\infty}^{\infty} \frac{f_i'(x_i)}{f_i(x_i)} f_i(x_i - \theta s_i)\,dx_i \tag{3.17}$$

Expanding in a Taylor series about $\theta = 0$ and simplifying yields

$$\mu_{iK} = \theta s_i^2 \, I(f_i) + \theta^3 \, J_1(\theta, s_i, f_i) \tag{3.18}$$

where $J_1(\theta, s_i, f_i)$ represents the remaining terms in the series. The variance under K may be treated similarly

$$\begin{aligned} \sigma_{iK}^2 &= E_K \left(- \frac{f_i'(x_i)}{f_i(x_i)} \, s_i \right)^2 - \mu_{iK}^2 \\ \sigma_{iK}^2 &= s_i^2 \, I(f_i) + \theta^2 \, J_2(\theta, s_i, f_i) \end{aligned} \tag{3.19}$$

Using the Gaussian approximation,

$$\alpha = \text{Prob}_H \left[T_{loc}(\underline{x}) > T \right] \approx 1 - \Phi \left[\frac{T - \sum_{i=1}^{N} \mu_{iH}}{\left(\sum_{i=1}^{N} \sigma_{iH}^2 \right)^{1/2}} \right] \tag{3.20}$$

$$T \approx \Phi^{-1}(1-\alpha) \left[\sum_{i=1}^{N} s_i^2 I(f_i) \right]^{1/2} \tag{3.21}$$

$$\beta(\theta) = \text{Prob}_K \left[T_{loc}(\underline{x}) > T \right] \approx 1 - \Phi \left[\frac{T - \sum_{i=1}^{N} \mu_{iK}}{\left(\sum_{i=1}^{N} \sigma_{iK}^2 \right)^{1/2}} \right] \tag{3.22}$$

Substituting the appropriate expressions gives

$$\beta(\theta) \approx 1-\Phi\left[\frac{\Phi^{-1}(1-\alpha) - k - \frac{\theta^4}{k}\sum_{i=1}^{N} J_1(\theta,s_i,f_i)}{\left(1 + \frac{\theta^4}{k^2}\sum_{i=1}^{N} J_2(\theta,s_i,f_i)\right)^{1/2}}\right] \tag{3.23}$$

where $k^2 = \theta^2 \sum_{i=1}^{N} s_i^2 I(f_i)$. In the limit as $\theta \to 0$ and $N \to \infty$ holding k constant, the power function is

$$\beta(\theta) = 1 - \Phi(\Phi^{-1}(1-\alpha) - k) \tag{3.24}$$

provided μ_{iK} and σ_{iK}^2 are bounded. Note that this equation for the asymptotic power function for the locally optimum detector is the same as the equation for the optimum detector [Eq. (3.13)]. This is not surprising since the Neyman-Pearson optimum detector approaches the normalized locally optimum detector as $N \to \infty$ and $\theta \to 0$.

The expression [Eq. (3.13) or (3.24)] for the asymptotic power of optimum and locally optimum detectors holds for all $f_i \in F$, where F is a class of symmetric densities with the following properties:

(1) f_i has infinite support with derivatives of all order on $(-\infty,\infty)$

(2) $f_i(\infty) = f_i(-\infty) = f_i'(\infty) = f_i'(-\infty) = 0$

(3) There exists positive constants m_2 and M_2 such that

(a) $0 < m_2 < I(f_i) < \infty$

and

(b) $$\int_{-\infty}^{\infty} \left|\frac{f_i'(x)}{f_i(x)}\right|^3 f_i(x)\,dx < M_2 < \infty$$

Condition (3) is required to satisfy sufficient conditions for the Central Limit Theorem (Appendix 3.2). In the special case where $s_i = 1$ for all i and the $\{N_i\}_1^N$ are identically distributed, that is $f_i = f$ for all i, Condition (3) may be replaced by the weaker condition:

(3') $I(f) < \infty$

The power function [Eq. (3.13) or equivalently Eq. (3.24)] is plotted in Fig. 3.1 for various values of α, the probability of false alarm. Note that $k^2 = \theta^2 \sum_{i=1}^{N} s_i^2 I(f_i)$ is basically a signal-to-noise power ratio since

$$\sum_{i=1}^{N} \theta^2 s_i^2 = \text{signal power parameter}$$

and

$$\left[\sum_{i=1}^{N} I(f_i) \right]^{-1} = \text{noise power parameter}$$

For the special case when $\sum_{i=1}^{N} s_i^2/N = 1$ and the $\{n_i\}_1^N$ have common density f, one obtains $k^2 = \theta^2 NI(f)$. Thus, the larger Fisher's Information is, the larger the generalized signal-to-noise ratio (k^2) will be, and the larger the probability of detection will become.

To illustrate the relationships between optimum detector structure, optimum detector performance, and Fisher's Information for the special case of $s_i = 1$ and $f_i = f$ for all i, two systems of densities will now be considered: a

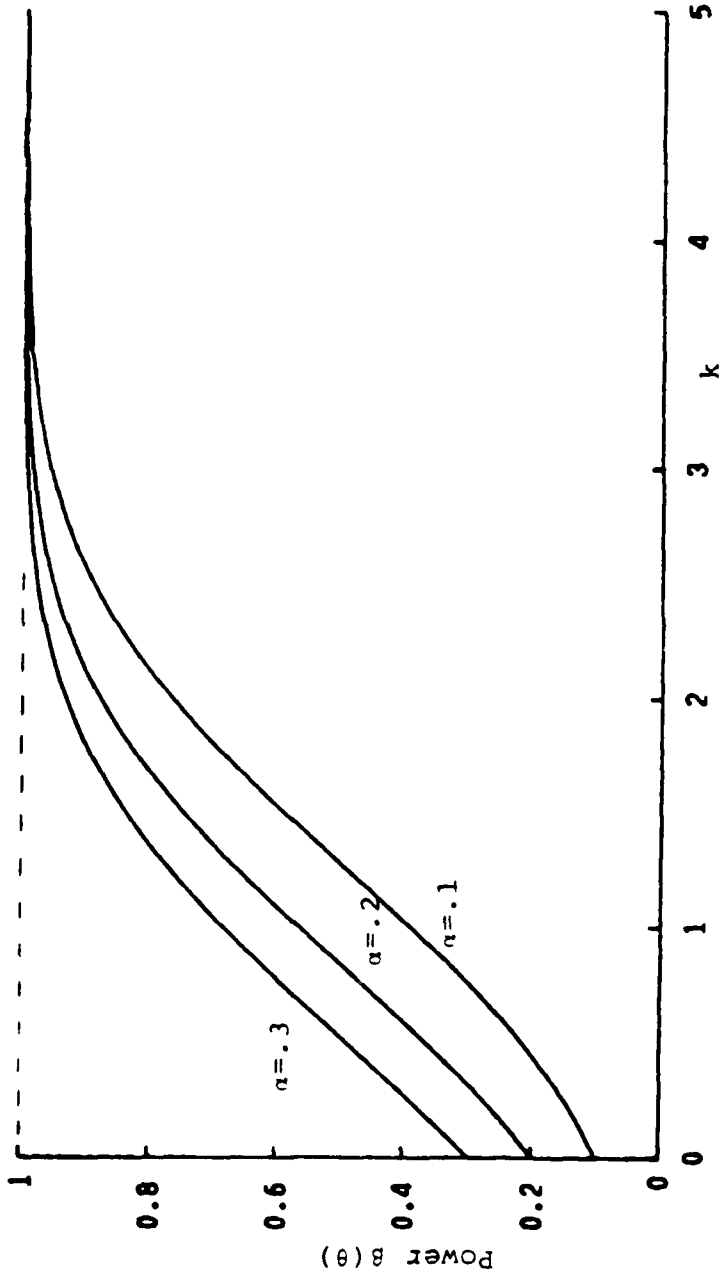


Fig. 3.1 - Power Curves

generalized Gaussian noise and the Johnson S_u System (see Chapter II - Section 3 for a description of these density systems, graphs of their densities and detectors, and plots of Fisher's Information vs. the system parameters).

The power curves for optimum detectors operating in generalized Gaussian noise for several values of the parameter c are given in Fig. 3.2. The power function formula calculated above only holds for $c > 1/2$. For $c \leq 1/2$ the sufficient conditions on the Central Limit Theorem are not satisfied [Condition (3') is violated]. The curves for small values of c, corresponding to large values for Fisher's Information (Fig. 2.12) are higher than those for values of c near 2. The $c = 1$ curve is the power function for the sign detector operating in Laplace noise. The $c = 2$ curve is for the linear detector in Gaussian noise. This set of curves illustrates the fact that it is easier to detect a signal in additive noise which has a large value for Fisher's Information than it is to detect a signal in noise with smaller values for Fisher's Information. One could consider Gaussian noise to be the most difficult noise environment in which to detect signals since Gaussian noise has minimum Fisher's Information among all fixed variance distributions and thus has a maximum value for the noise power parameter among all densities of a given r.m.s. noise power.

Fig. 3.3 displays the power curves for optimum detectors operating in noise from the Johnson S_u System. Small values of δ, corresponding to large values for Fisher's Information (Fig. 2.16), yield higher curves than large values of δ. As δ increases, the curves rapidly approach the power curve for the linear detector in Gaussian noise.

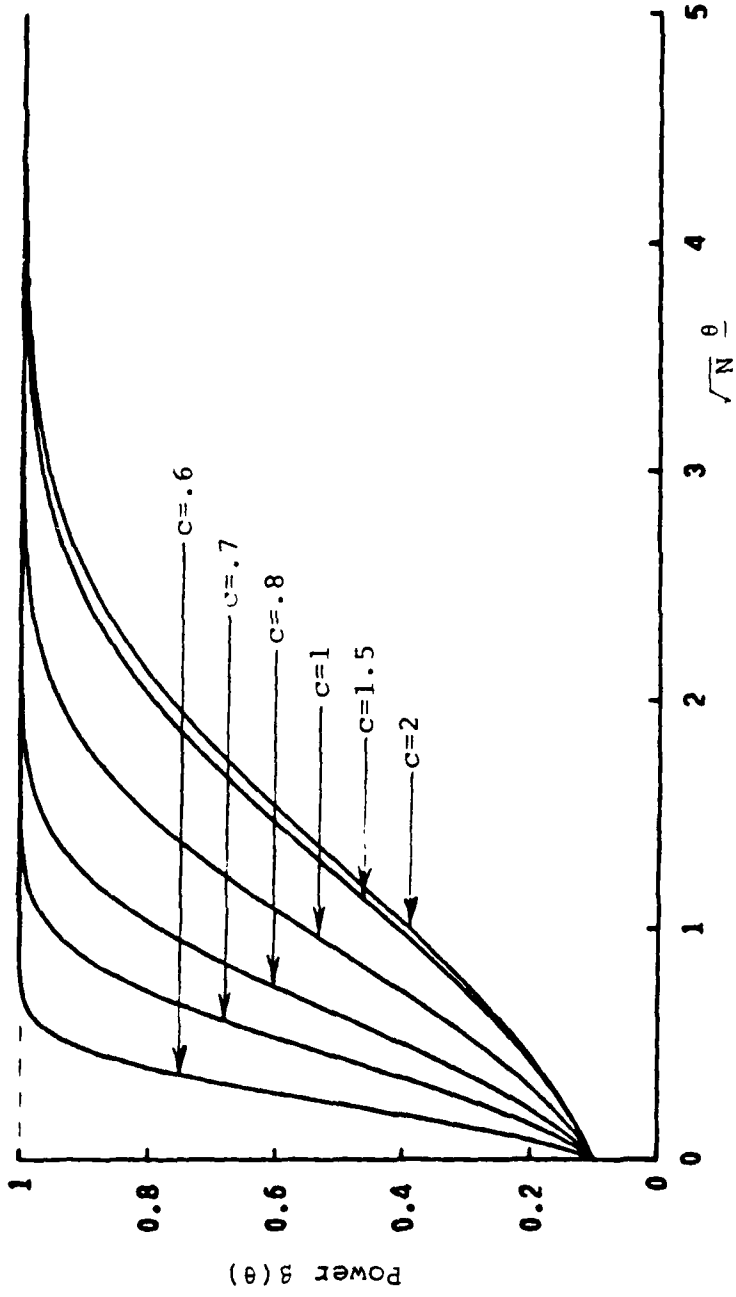


Fig. 3.2 Generalized Gaussian Noise: Power Curves, $\alpha = 0.1$

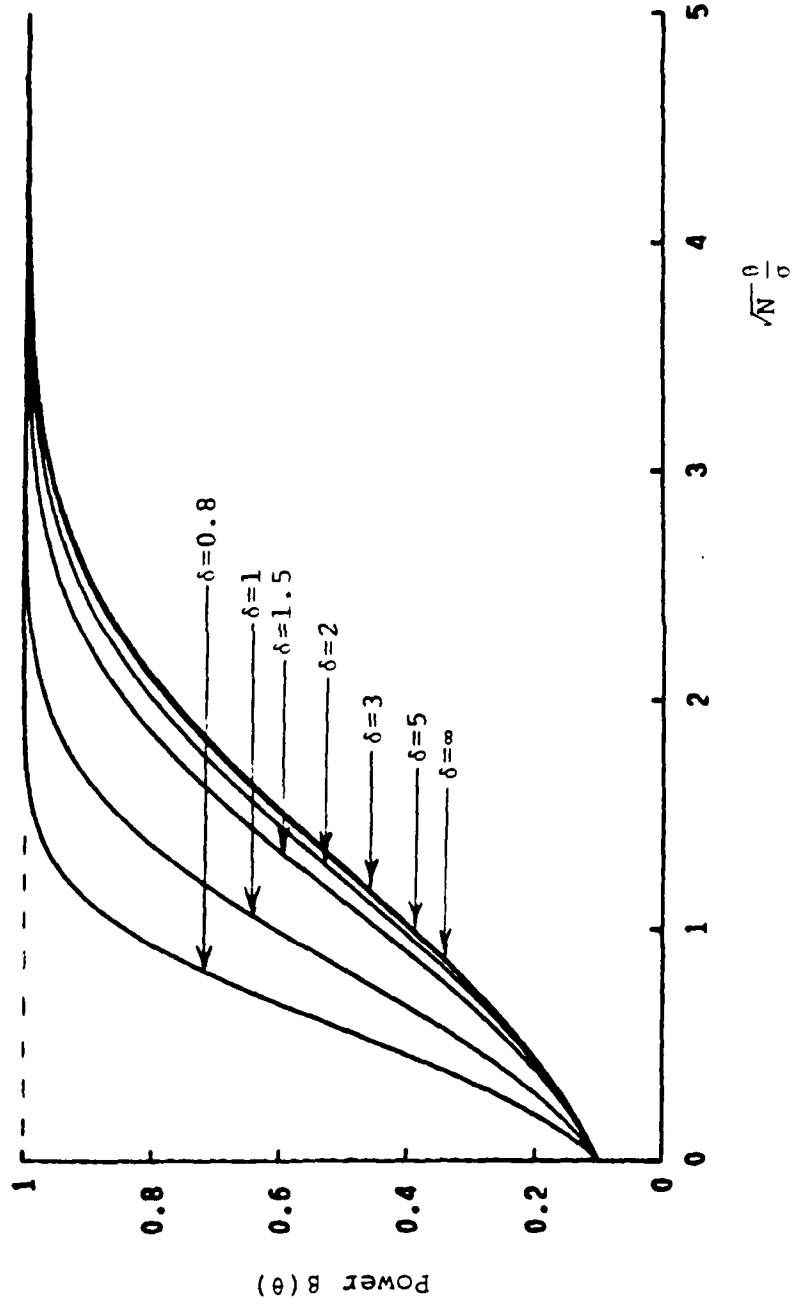


Fig. 3.3 - Johnson's S_u System: Power Curves, α = 0.1

All three of these graphs serve to illustrate the fact that Fisher's Information can be used as a measure of the difficulty of detecting an additive signal in noise. Non-Gaussian noise, with large values for Fisher's Information, increases k^2 thus providing higher power at a given level than is possible for the Gaussian distribution. Impulsive noise, with its characteristic heavy tails, has a high value for Fisher's Information. The optimum detector for this noise will perform far better than the linear detector would in Gaussian noise. This is reasonable since limiters or blankers can be used to suppress the heavy tails which in effect will decrease the noise power without decreasing the signal power. Hence, the increased structure of non-Gaussian noise, if properly considered, can be a great asset. Simulation results are presented in Chapter IV which support these statements.

Although the expression for the asymptotic power function was derived assuming independent observations, the Central Limit Theorem holds under the weaker condition of strong mixing [8]. The expression will then still be valid; however, the detector in question will no longer be optimum and also the value for k will change since the dependency structure will introduce cross terms in the computation of σ_{iH}^2 and σ_{iK}^2.

Section 2: The Relationship Between Efficacy and the Asymptotic Performance of Nonlinear Detectors

In Section 1 it has been shown that optimum detectors for densities with high values for Fisher's Information will perform

better than optimum detectors for densities with low values for Fisher's Information. Therefore, Fisher's Information can be used to rank the densities within a class of densities to determine which density provides the worst environment for detection. However, in order to prove that the pair (f_o, D_o) is a saddlepoint (f_o minimizes Fisher's Information over the class and D_o is the locally optimum detector for f_o), the results of Section 1 must be generalized to account for all nonlinear detectors.

Consider a special case of the problem treated in Section 1 of this chapter.

$$\begin{aligned} H:&\quad x_i = n_i \\ K:&\quad x_i = n_i + \theta \end{aligned} \qquad \theta > 0,\ i=1,\ldots,N \tag{3.25}$$

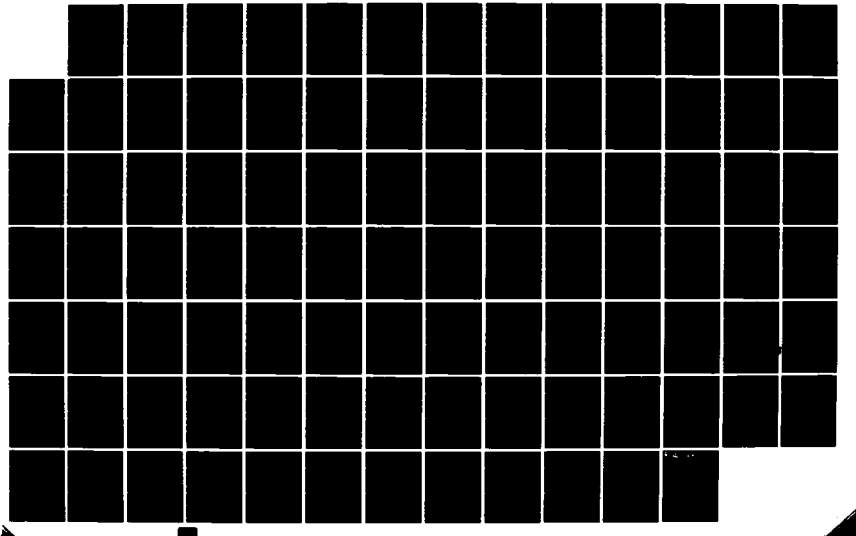
The known signal θ is assumed constant. The $\{n_i\}_1^N$ are assumed to be realizations of the independent and identically distributed random variables $\{N_i\}_1^N$. The noise is distributed according to f, where $f \in F$ a class of zero mean symmetric densities having finite Fisher's Information.

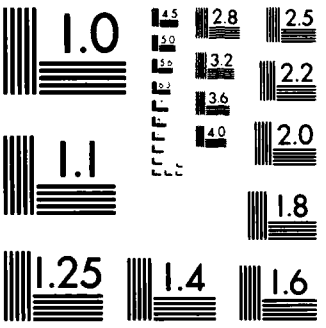
The canonical nonlinear detector for this problem can be written

$$T(\underline{x}) = \sum_{i=1}^{N} g(x_i) \quad \begin{cases} > T & \text{decide K} \\ < T & \text{decide H} \end{cases} \tag{3.26}$$

where $g(\cdot)$ is a memoryless nonlinearity. The threshold T is chosen to achieve the desired probability of false alarm. Through the proper selection of the nonlinearity, this equation

AD-A130 393 THE DETECTION OF SIGNALS IN IMPULSIVE NOISE(U) 2/2
PRINCETON UNIV NJ INFORMATION SCIENCES AND SYSTEMS LAB
E J MODUGNO ET AL. JUN 83 TR-13 N00014-81-K-0146
UNCLASSIFIED F/G 9/4 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

can represent both the optimum and locally optimum detectors for the problem given by Eq. (3.25). Note that this detector is both memoryless and non-time-varying. This is because the signal is constant and the noise samples are assumed to be independent and identically distributed.

As in Section 1, the Central Limit Theorem can be used to calculate the asymptotic power function for the detector given by Eq. (3.26). When the signal is absent, the mean and variance of the single sample test statistic are

$$\mu_H = E_H[g(x)] = \int_{-\infty}^{\infty} g(x)f(x)\,dx$$

$$\sigma_H^2 = E_H[g^2(x)] - \mu_H^2 = \int_{-\infty}^{\infty} g^2(x)f(x)\,dx - \left(\int_{-\infty}^{\infty} g(x)f(x)\,dx\right)^2 \tag{3.27}$$

When the signal is present,

$$\mu_K = E_K[g(x)] = \int_{-\infty}^{\infty} g(x)f(x-\theta)\,dx \tag{3.28}$$

By a change of variables in the integrand, this becomes

$$\mu_K = \int_{-\infty}^{\infty} g(x+\theta)f(x)\,dx \tag{3.29}$$

Also,

$$\sigma_K^2 = E_K[g^2(x)] - \mu_K^2$$

$$\sigma_K^2 = \int_{-\infty}^{\infty} g^2(x+\theta)f(x)\,dx - \left(\int_{-\infty}^{\infty} g(x+\theta)f(x)\,dx\right)^2 \tag{3.30}$$

Using the Gaussian approximation provided by the Central Limit Theorem, the threshold is chosen to fix α:

$$\alpha = \text{Prob}_H \left[T(\underline{x}) > T \right] \approx 1 - \Phi\left[\frac{T - N\mu_H}{\left(N\sigma_H^2 \right)^{1/2}} \right] \tag{3.31}$$

$$T = \Phi^{-1}(1-\alpha)\sqrt{N}\,\sigma_H + N\mu_H \tag{3.32}$$

The power function is then:

$$\beta = \text{Prob}_K \left[T(\underline{x}) > T \right] \approx 1 - \Phi\left[\frac{T - N\mu_K}{\left(N\sigma_K^2 \right)^{1/2}} \right] \tag{3.33}$$

$$\beta \approx 1 - \Phi\left(\Phi^{-1}(1-\alpha)\,\frac{\sigma_H}{\sigma_K} - \sqrt{N}\,\frac{\mu_K - \mu_H}{\sigma_K} \right) \tag{3.34}$$

Now,

$$\left(\frac{\sigma_H}{\sigma_K} \right)^2 = \frac{\int_{-\infty}^{\infty} g^2(x) f(x)\,dx - \left(\int_{-\infty}^{\infty} g(x) f(x)\,dx \right)^2}{\int_{-\infty}^{\infty} g^2(x+\theta) f(x)\,dx - \left(\int_{-\infty}^{\infty} g(x+\theta) f(x)\,dx \right)^2} \tag{3.35}$$

Expanding $g(x+\theta)$ and $g^2(x+\theta)$ in a Taylor series, it can be easily seen that

$$\lim_{\theta \to 0} \left(\frac{\sigma_H}{\sigma_K} \right)^2 = 1 \tag{3.36}$$

Also,

$$\sqrt{N}\,\frac{\mu_K - \mu_H}{\sigma_K} = \sqrt{N}\,\frac{\int_{-\infty}^{\infty} g(x+\theta) f(x)\,dx - \int_{-\infty}^{\infty} g(x) f(x)\,dx}{\left[\int_{-\infty}^{\infty} g^2(x+\theta) f(x)\,dx - \left(\int_{-\infty}^{\infty} g(x+\theta) f(x)\,dx \right)^2 \right]^{1/2}} \tag{3.37}$$

In the limit as $\theta \to 0$ and $N \to \infty$ holding $\sqrt{N}\,\theta = k_o$ = a constant,

$$\sqrt{N}\,\frac{\mu_K - \mu_H}{\sigma_K} \;\to\; \sqrt{N}\,\theta\sqrt{E_g(f)} \tag{3.38}$$

where

$$E_g(f) = \frac{\left[\int_{-\infty}^{\infty} g'(x) f(x)\, dx\right]^2}{\int_{-\infty}^{\infty} g^2(x) f(x)\, dx - \left[\int_{-\infty}^{\infty} g(x) f(x)\, dx\right]^2} \tag{3.39}$$

is the efficacy of the nonlinear detector (see Appendix 3.3). The asymptotic power function can now be written as

$$\beta = 1 - \Phi(\Phi^{-1}(1-\alpha) - \sqrt{N}\,\theta\,\sqrt{E_g(f)}\,) \tag{3.40}$$

Efficacy thus has a monotone relationship with the asymptotic performance of nonlinear detectors. High values for efficacy correspond to higher power at a given signal strength than do low values. Note that for $g(x) = -f'(x)/f(x)$ (the locally optimum detector), one finds $E_g(f) = I(f)$ (Fisher's Information), and Eq. (3.40) reduces to the result from Section 1 [Eq. (3.24)].

Section 3: The Saddlepoint for the Minimax Problem

The results of Section 2 may be used to prove that the pair (f_o, D_o), where f_o minimizes Fisher's Information over the

given class of densities and D_o is the locally optimum detector for density f_o, is a saddlepoint for the minimax detection problem previously described. Let f be contained in F, a convex set of densities, and let g be any antisymmetric detector nonlinearity. Also let g_o be the locally optimum nonlinearity for density f_o (i.e. $g_o(x) = -f_o'(x)/f_o(x)$). If f_o and g_o satisfy

$$E_g(f_o) \leq E_{g_o}(f_o) \leq E_{g_o}(f) \tag{3.41}$$

for all $f \in F$ and all g antisymmetric, then (f_o, D_o) is a saddlepoint. That is, any g other than g_o will decrease the efficacy (and thus the power function) when f_o is the noise. Also, any other noise f will increase the efficacy when g_o is used for detection. Thus f_o is the "worst density" and g_o is the best detector for this density. This is the minimax solution. The detector is chosen for minimum error using the density which maximizes error. This problem is related to the minimax estimation problem of [9].

The proof of Eq. (3.41) is similar to one used by Huber [10] in his work on robust estimation. His equation for the asymptotic variance of an M-estimate is similar to the equation for efficacy.

For g(x) antisymmetric, efficacy can be written

$$E_g(f) = \frac{\left[\int_{-\infty}^{\infty} g'(x) f(x)\, dx\right]^2}{\int_{-\infty}^{\infty} g^2(x) f(x)\, dx} \tag{3.42}$$

The relationship $E_g(f_o) \leq E_{g_o}(f_o)$ follows from the fact that the nonlinear detector which maximizes efficacy is the locally

optimum detector [11]. To prove $E_{g_o}(f_o) \leq E_{g_o}(f)$, first note that $E_{g_o}(f)$ is a convex function of f (See Appendix 3.4). Then let $f_t = (1-t)f_o + tf_1$ where $f_1 \in F$. Now,

$$\frac{d}{dt} E_{g_o}(f_t)\Big|_{t=0} \geq 0 \tag{3.43}$$

if and only if $E_{g_o}(f)$ is increasing in all directions from f_o. Which, by the convexity of $E_{g_o}(f)$ on the convex set of densities F, means that f_o must minimize $E_{g_o}(f)$. By a straightforward computation,

$$\frac{d}{dt} E_{g_o}(f_t)\Big|_{t=0} = \frac{d}{dt}\left(\frac{\left[(1-t)\int_{-\infty}^{\infty} g_o'(x) f_o(x)\,dx + t\int_{-\infty}^{\infty} g_o'(x) f_1(x)\,dx\right]^2}{(1-t)\int_{-\infty}^{\infty} g_o^2(x) f_o(x)\,dx + t\int_{-\infty}^{\infty} g_o^2(x) f_1(x)\,dx}\right)\Bigg|_{t=0}$$

$$= \frac{2\int_{-\infty}^{\infty} g_o'(x) f_o(x)\,dx}{\int_{-\infty}^{\infty} g_o^2(x) f_o(x)\,dx} \int_{-\infty}^{\infty} g_o'(x)\left[f_1(x) - f_o(x)\right]dx \tag{3.44}$$

$$- \frac{\left[\int_{-\infty}^{\infty} g_o'(x) f_o(x)\,dx\right]^2}{\left[\int_{-\infty}^{\infty} g_o^2(x) f_o(x)\,dx\right]^2} \int_{-\infty}^{\infty} g_o^2(x)\left[f_1(x) - f_o(x)\right]dx$$

To simplify this expression, note that

$$\int_{-\infty}^{\infty} g_o'(x) f_o(x)\,dx = \int_{-\infty}^{\infty} (-1)\left[\frac{f_o''(x)}{f_o(x)} - \left(\frac{f_o'(x)}{f_o(x)}\right)^2\right] f_o(x)\,dx \tag{3.45}$$

$$= I(f_o)$$

and

$$\int_{-\infty}^{\infty} g_o^{\,2}(x) f_o(x)\, dx = \int_{-\infty}^{\infty} \left(- \frac{f_o'(x)}{f_o(x)} \right)^2 f_o(x)\, dx = I(f_o) \tag{3.46}$$

where $I(f_o)$ is Fisher's Information for density f_o. Thus,

$$\frac{d}{dt}\, E_{g_o}(f_t)\Big|_{t=0} = \int_{-\infty}^{\infty} \left[2g_o'(x) - g_o^{\,2}(x) \right] \left[f_1(x) - f_o(x) \right] dx \tag{3.47}$$

Now this equation is equivalent to the condition necessary for the minimization of Fisher's Information. That is,

$$\frac{d}{dt}\, I(f_t)\Big|_{t=0} = \frac{d}{dt} \int_{-\infty}^{\infty} \frac{\left[(1-t) f_o'(x) + t f_1'(x) \right]^2}{(1-t) f_o(x) + t f_1(x)}\, dx \Bigg|_{t=0} \tag{3.48}$$

$$= -2 \int_{-\infty}^{\infty} g_o(x) \left[f_1'(x) - f_o'(x) \right] dx - \int_{-\infty}^{\infty} g_o^{\,2}(x) \left[f_1(x) - f_o(x) \right] dx$$

Integrating the first integral by parts leaves

$$\frac{d}{dt}\, I(f_t)\Big|_{t=0} = \int_{-\infty}^{\infty} \left[2g_o'(x) - g_o^{\,2}(x) \right] \left[f_1(x) - f_o(x) \right] dx \tag{3.49}$$

So I(f) will be minimized at $f = f_o$ if and only if Eq. (3.49) is nonnegative. Thus, $E_{g_o}(f)$ is minimized at f_o if and only if I(f) is minimized at f_o, which occurs if and only if

$$\int_{-\infty}^{\infty} [2g_o'(x) - g_o^2(x)][f_1(x) - f_o(x)]\, dx \geq 0 \qquad (3.50)$$

for all $f_1 \in F$. Therefore, $E_{g_o}(f_o) \leq E_{g_o}(f)$ when f_o minimizes Fisher's Information over the class F. Eq. (3.41) holds and the pair (f_o, D_o) is a saddlepoint.

Section 4: Minimax Detectors - Some Examples

Sections 1, 2, and 3 have established the fact that given a convex set of densities F, the minimax detector for F is the locally optimum detector D_o for the density f_o which minimizes Fisher's Information over the class. Since $I(f) = E_{(-f'/f)}(f)$ is a convex function in f (see Appendix 3.4) there exists a density with minimum Fisher's Information. Huber [3] has shown that this density is also unique. In this section several different classes of densities are considered and conditions are established for the minimax detectors.

First M Moments Fixed

Consider the class consisting of all densities which have the values $\{\mu_m\}_1^M$ for the first M moments. That is, a density f(x) in this class must satisfy

$$\int_{-\infty}^{\infty} x^m f(x)\, dx = \mu_m \qquad m = 0,1,2,\ldots,M \tag{3.51}$$

where $\mu_o = 1$. The density with minimum Fisher's Information in this class may be found using the Euler-Lagrange equation

$$\frac{\partial L}{\partial f} - \frac{d}{dx}\left(\frac{\partial L}{\partial f'}\right) = 0 \tag{3.52}$$

with

$$L(x,f,f') = \left(\frac{f'(x)}{f(x)}\right)^2 f(x) + \sum_{m=0}^{M} \lambda_m x^m f(x)$$

A straightforward calculation yields the following second order nonlinear differential equation

$$\left(\frac{f'(x)}{f(x)}\right)^2 - \frac{2f''(x)}{f(x)} + \sum_{m=0}^{M} \lambda_m x^m = 0 \tag{3.53}$$

The density which satisfies Eq. (3.53) and the constraints given by Eq. (3.51) is the density with minimum Fisher's Information. The locally optimum detector for this density is the asymptotic minimax detector for the class.

As an example, consider the case $M = 2$ with $\mu_1 = 0$ and $\mu_2 = \sigma^2$ (i.e., all densities with fixed variance σ^2). Eq. (3.53) becomes

$$\left(\frac{f'(x)}{f(x)}\right)^2 - 2\,\frac{f''(x)}{f(x)} + \lambda_o + \lambda_2 x^2 = 0 \tag{3.54}$$

with the constraints

$$\int_{-\infty}^{\infty} f(x)\, dx = 1$$

$$\int_{-\infty}^{\infty} x^2 f(x)\, dx = \sigma^2$$

These equations are satisfied by the Gaussian distribution. The linear detector is the minimax detector for the class. It guarantees the lower bound of $1/\sigma^2$ for the efficacy. The following table displays the efficacy for both the linear detector and the locally optimum detector for some noise f_1.

Efficacy

	Gaussian noise (σ^2=1)	non-Gaussian noise f_1 (σ^2=1)
linear detector	1	1
locally optimum detector ($-f_1'/f_1$)	some amount ≤ 1	$I(f_1) \geq 1$

Use of the linear detector guarantees an efficacy of unity. If another detector is used, the efficacy for Gaussian noise will be less than unity. However, when the noise f_1 for which the detector was designed is present, the efficacy will be given by Fisher's Information (see Section 2) and could be much greater than unity. If one either has some prior knowledge of the noise or is willing to accept a decreased performance for Gaussian noise then one could do far better than the linear detector, even

though it is the most robust (in the minimax sense) for this class. A conclusion which can be drawn is that this class of fixed moment densities may not be very realistic.

A Class of Mixtures

Perhaps a more useful class of densites is given by the mixture

$$f(x) = (1 - \varepsilon)g(x) + \varepsilon h(x) \qquad , 0 < \varepsilon < 1 \tag{3.55}$$

where $g(x)$ is some known symmetric density (with $-\log g(x)$ convex) and $h(x)$ is an unknown contaminant which is assumed symmetric. Huber [3] has shown that the density $f_o(x)$ with minimum Fisher's Information in this class is given by

$$f_o(x) = \begin{cases} (1 - \varepsilon)g(x) & , |x| \leq a \\ (1 - \varepsilon)g(a)e^{-k(|x|-a)} & , |x| > a \end{cases} \tag{3.56}$$

where $[-a,a]$ is an interval such that $|g'(x)/g(x)| \leq k$ and

$$\int_{-a}^{a} g(x)\, dx + \frac{2g(a)}{k} = \frac{1}{1 - \varepsilon} \tag{3.57}$$

A proof of this result is given in Appendix 3.5. When $g(x)$ is the Gaussian distribution, the locally optimum detector for density $f_o(x)$ is an amplifier limiter. It's performance has been studied by Martin and Schwartz [12]. Kassam and Thomas [13]

have studied the various limiters which arise when g(x) is a member of the generalized Gaussian distribution defined in Chapter III - Section 3.

Other Possibilities

Huber [14] has developed a technique to determine the distribution with minimum Fisher's Information which will fit a set of measured points. Given $k \geq 2$ points, Huber has found that there is a unique function which has a minimum value for Fisher's Information among all distributions which pass through the points. This function can be obtained by spline interpolation. Let $F(\xi_i) = t_i$, $i = 1,\ldots,k$ be the k points. Assume that $\xi_o = -\infty$, $t_o = 0$ and $\xi_{k+1} = \infty$, $t_{k+1} = 1$. Then the distribution $F_o(x)$ with minimum Fisher's Information which passes through these points must satisfy the following four conditions:

(1) $F_o(\xi_i) = t_i$, $i = 0,1,\ldots,k+1$

(2) F_o is two times continuously differentiable

(3) $f_o(x) = F_o'(x) > 0$ except over intervals $[\xi_i, \xi_{i+1}]$ where $t_i = t_{i+1}$ (in such cases $F_o'(x) = 0$)

(4) On each interval $(\sqrt{f_o})''/\sqrt{f_o} = \lambda_i$ = a constant

Also, the value for Fisher's Information will be

$$I(f_o) = -4 \sum_{i=0}^{k} (t_{i+1} - t_i)\lambda_i \qquad (3.58)$$

This result could be very useful for detection systems. Initially measurements would be made and the density which satisfies conditions (1) through (4) would be found. The optimum detector for this density should perform well for the actual noise.

Two other classes of densities in which the member with minimum Fisher's Information has been found are: The p-point class and Huber's ϵ-normal class. The p-point class [4] is defined as the set of all symmetric distributions which are continuous at $\pm$ a and which satisfy $\int_{-a}^{a} f(x)\,dx = p$ for some fixed a and p. Huber's ϵ-normal class [3,10] is the set of all distributions which differ by at most ϵ in Kolmogorov distance from the standard normal cummulative distribution function: $\sup_x |F(x) - \Phi(x)| \leq \epsilon$. Both of these classes may be used for minimax detection.

Appendix 3.1

The Relationship between the Efficacy of the Locally Optimum Detector and Fisher's Information

The efficacy of a detector is given by Eq. (1.15) as,

$$E_T = \lim_{N\to\infty} \frac{\left[\frac{\partial}{\partial\theta} E_K[T(\underline{x})] \mid_{\theta\to 0}\right]^2}{N \, \mathrm{Var}_H(T(\underline{x}))}$$

For the locally optimum detector $T(\underline{x}) = \sum_{i=1}^{N} \frac{-f_i'(x_i)}{f_i(x_i)} s_i$

From Eqs. (3.16) and (3.18),

$$E_K[T(\underline{x})] = \sum_{i=1}^{N} \mu_{iK} = \theta \sum_{i=1}^{N} s_i^2 I(f_i) + \theta^3 \sum_{i=1}^{N} J_1(\theta, s_i, f_i)$$

$$\mathrm{Var}_H[T(\underline{x})] = \sum_{i=1}^{N} \sigma_{iH}^2 = \sum_{i=1}^{N} s_i^2 I(f_i)$$

Thus,

$$E_{loc} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} s_i^2 I(f_i)$$

For the special case of $\frac{1}{N} \sum_{i=1}^{N} s_i^2 = 1$ and $f_i = f$ for all i,

$$E_{loc} = I(f)$$

Appendix 3.2

Satisfying the Sufficient Conditions for the Central Limit Theorem

For the Central Limit Theorem to hold, it is sufficient that:

$$(1)\ \sigma_i^2 < \infty$$

$$(2)\ \sigma_i^2 > m > 0$$

$$(3)\ E\{|g_i(x_i)-\mu_i|^3\} < M < \infty$$

for all i, under both H and K. In the special case of $f_i=f$ and $s_i=1$ for all i, Condition (1) alone is sufficient.

First consider the Neyman-Pearson detector.

$$\sigma_{iH}^2 = \sigma_{iK}^2 = \theta^2 s_i^2 I(f_i) + \theta^4 K_2(\theta, s_i, f_i)$$

For sufficiently small θ, Condition (1) is satisfied by requiring $|s_i| < \infty$ and $I(f_i) < \infty$. Condition (2) is met if $|s_i| > m_1 > 0$ and $I(f_i) > m_2 > 0$. Utilizing a Taylor series, Condition (3) can be simplified

$$E\left\{\left|\log \frac{f_i(x_i-\theta s_i)}{f_i(x_i)} - \mu_i\right|^3\right\} \leq E\left\{\left|\log \frac{f_i(x_i-\theta s_i)}{f_i(x_i)}\right|^3 + |\mu_i|^3\right\}$$

$$\leq \theta^3 |s_i|^3 \int_{-\infty}^{\infty} \left|\frac{f_i'(x_i)}{f_i(x_i)}\right|^3 f_i(x_i)\, dx_i + \{\text{higher order terms}\}$$

Thus, for sufficiently small θ, condition (3) can be satisfied, under both H and K, by requiring $|s_i| < M_1 < \infty$ and

$$\int_{-\infty}^{\infty} \left|\frac{f_i'(x_i)}{f_i(x_i)}\right|^3 f_i(x_i)\, dx_i < M_2 < \infty$$

For the locally optimum detector,

$$\sigma_{iH}^2 = s_i^2 I(f_i)$$

and

$$\sigma_{iK}^2 = s_i^2 I(f_i) + \theta^2 J_2(\theta, s_i, f_i).$$

For sufficiently small θ, Conditions (1) and (2) can be satisfied by requiring $0 < m_1 < |s_i| < \infty$ and $0 < m_2 < I(f_i) < \infty$ just as in the Neyman-Pearson case. The left hand side of Condition (3) under both H and K can be written

$$E\left\{ \left| \frac{-f_i'(x_i)}{f_i(x_i)} s_i - \mu_i \right|^3 \right\} \leq |s_i|^3 E\left\{ \left| \frac{f_i'(x_i)}{f_i(x_i)} \right|^3 \right\} + |\mu_i|^3$$

$$\leq |s_i|^3 \int_{-\infty}^{\infty} \left| \frac{f_i'(x_i)}{f_i(x_i)} \right|^3 f_i(x_i)\, dx_i + \{\text{higher order terms}\}$$

Thus, Condition (3) can be satisfied if $|s_i| < M_1 < \infty$ and

$$\int_{-\infty}^{\infty} \left| \frac{f_i'(x_i)}{f_i(x_i)} \right|^3 f_i(x_i)\, dx_i < M_2 < \infty$$

Appendix 3.3

The Efficacy of the Nonlinear Detector

The efficacy of a detector is given by Eq. (1.15) as,

$$E_T = \lim_{N\to\infty} \frac{\left[\frac{d}{d\theta} E_K[T(\underline{x})]\Big|_{\theta\to 0}\right]^2}{N\mathrm{Var}_H[T(\underline{x})]}$$

Using the nonlinear detector of Eq. (3.26),

$$T(\underline{x}) = \sum_{i=1}^{N} g(x_i)$$

$$E_K[T(\underline{x})] = NE_K[g(x)] = N\int_{-\infty}^{\infty} g(x+\theta) f(x)\, dx$$

$$\frac{d}{d\theta} E_K[T(\underline{x})]\Big|_{\theta\to 0} = N\int_{-\infty}^{\infty} \frac{d}{d\theta} g(x+\theta)\Big|_{\theta\to 0} f(x)\, dx$$

$$= N\int_{-\infty}^{\infty} g'(x) f(x)\, dx$$

Also,

$$\mathrm{Var}_H[T(\underline{x})] = N\mathrm{Var}_H[g(\underline{x})]$$

$$= N\left[\int_{-\infty}^{\infty} g^2(x) f(x)\, dx - \left(\int_{-\infty}^{\infty} g(x) f(x)\, dx\right)^2\right]$$

Thus, for the nonlinear detector,

$$E_T = E_g(f) = \frac{\left[\int_{-\infty}^{\infty} g'(x) f(x)\, dx\right]^2}{\int_{-\infty}^{\infty} g^2(x) f(x)\, dx - \left[\int_{-\infty}^{\infty} g(x) f(x)\, dx\right]^2}$$

For the special case $g = -\frac{f'(x)}{f(x)}$, $E_g(f) = I(f)$ as in Appendix 3.1.

Appendix 3.4

A Proof that Efficacy is a Convex Function of f when g is Antisymmetric

Efficacy, for g(x) antisymmetric, is given by Eq. (3.42) as,

$$E_g(f) = \frac{\left[\int_{-\infty}^{\infty} g'(x) f(x)\ dx\right]^2}{\int_{-\infty}^{\infty} g^2(x) f(x)\ dx}$$

Let

$$U(f) = \int_{-\infty}^{\infty} g'(x) f(x)\ dx$$

and

$$V(f) = \int_{-\infty}^{\infty} g^2(x) f(x)\ dx$$

Note that $U''(f) = V''(f) = 0$ and $V(f) > 0$ for all but degenerate cases. Thus,

$$E_g(f) = W(f) = \frac{[U(f)]^2}{V(f)}$$

is a convex function of f because

$$W''(f) = \frac{2[U'V - UV']^2}{V^3} \geq 0$$

In particular $I(f) = E_{-\frac{f'}{f}}(f)$ is convex because the locally optimum detector is antisymmetric.

Appendix 3.5

A Proof that $f_o(x)$ of Eq. (3.56) Minimizes Fisher's Information for the Mixture Class Specified by Eq. (3.55)

This proof is similar to one given by Huber [10]. First $f_o(x)$ is proven to belong to the mixture class. Then it is proven that $f_o(x)$ minimizes Fisher's Information over the class. From Eqs. (3.56) and (3.57)

$$f_o(x) = \begin{cases} (1-\varepsilon)g(x) & ,|x| \leq a \\ (1-\varepsilon)g(a)e^{-k(|x|-a)} & ,|x| > a \end{cases}$$

where $[-a,a]$ is an interval such that $\left|\frac{g'(x)}{g(x)}\right| \leq k$ and

$$\int_{-a}^{a} g(x)\,dx + \frac{2g(a)}{k} = \frac{1}{(1-\varepsilon)}$$

The class is defined as all densities of the form

$$f(x) = (1-\varepsilon)g(x) + \varepsilon h(x)$$

where $g(x)$ and $h(x)$ are symmetric densities, $0 \leq \varepsilon \leq 1$, and $-\log g(x)$ is convex. The density $f_o(x)$ is an element of the class if

$$h_o(x) = \frac{1}{\varepsilon} f_o(x) - \frac{(1-\varepsilon)}{\varepsilon} g(x)$$

$$= \begin{cases} 0 & ,|x| \leq a \\ \frac{(1-\varepsilon)}{\varepsilon} g(a)e^{-k(|x|-a)} - \frac{(1-\varepsilon)}{\varepsilon} g(x) & ,|x| > a \end{cases}$$

is a density. To show $h_o(x) \geq 0$, note that since $-\log g(x)$ is convex, it lies above its tangents at $\pm a$.

$$-\log g(x) \geq -\log g(a) + \left(-\frac{g'(a)}{g(a)}\right)(x - a) \quad , x > a$$

$$-\log g(x) \geq -\log g(-a) + \left(-\frac{g'(-a)}{g(-a)}\right)(x + a) \quad , x < -a$$

Also, $-\frac{g'(a)}{g(a)} = k$ and $-\frac{g'(-a)}{g(-a)} = -k$. So that,

$$-\log g(x) \geq -\log g(a) + k(|x| - a)$$

Thus,

$$g(x) \leq g(a)e^{-k(|x|-a)} \quad , |x| > a$$

and

$$\frac{(1 - \varepsilon)}{\varepsilon} g(a)e^{-k(|x|-a)} - \frac{(1 - \varepsilon)}{\varepsilon} g(x) \geq 0 \quad \text{for } |x| > a$$

Hence $h_o(x) \geq 0$. Also,

$$\int_{-\infty}^{\infty} h_o(x)\, dx = \frac{(1 - \varepsilon)}{\varepsilon} g(a) \int_{|x|>a} e^{-k(|x|-a)}\, dx - \frac{(1 - \varepsilon)}{\varepsilon} \int_{|x|>a} g(x)\, dx$$

$$= \frac{(1 - \varepsilon)}{\varepsilon} g(a)\, \frac{2}{k} - \frac{(1 - \varepsilon)}{\varepsilon} \left[1 - \int_{-a}^{a} g(x)\, dx\right]$$

$$= \frac{(1 - \varepsilon)}{\varepsilon} \left[\frac{2g(a)}{k} + \int_{-a}^{a} g(x)\, dx\right] - \frac{(1 - \varepsilon)}{\varepsilon}$$

$$= \frac{1}{\varepsilon} - \frac{(1 - \varepsilon)}{\varepsilon} = 1$$

Thus $h_o(x)$ is a density and $f_o(x)$ is contained in the mixture class of interest.

From Eq. (3.50), $f_o(x)$ minimizes Fisher's Information if and only if

$$J = \int_{-\infty}^{\infty} [2g_o'(x) - g_o^2(x)]\ [f_1(x) - f_o(x)]\ dx \geq 0$$

for all $f_1(x)$ in the class. Now,

$$g_o(x) = \begin{cases} -\frac{g'(x)}{g(x)} & ,|x| \leq a \\ k\ \mathrm{sgn}(x) & ,|x| > a \end{cases}$$

$$g_o'(x) = \begin{cases} \left(-\frac{g'(x)}{g(x)}\right)' & ,|x| \leq a \\ 0 & ,|x| > a \end{cases}$$

$$-\log g(x) \text{ convex} \Rightarrow \left(-\frac{g'(x)}{g(x)}\right)' > 0 \Rightarrow g_o'(x) \geq 0$$

Also, $\left|\frac{g'(x)}{g(x)}\right| \leq k$ for $|x| \leq a$. Thus, $k^2 - g_o^2(x) \geq 0$ and $2g_o'(x) - g_o^2(x) + k^2 \geq 0$ for $|x| \leq a$. For $|x| > a$, $g_o'(x) = 0$ and $g_o^2(x) = k^2$; thus $2g_o'(x) - g_o^2(x) + k^2 = 0$. Then, adding and subtracting k^2 from inside the integrand

$$J = \int_{-a}^{a} [2g_o'(x) - g_o^2(x) + k^2][f_1(x) - f_o(x)]dx - k^2\int_{-\infty}^{\infty} [f_1(x) - f_o(x)]dx$$

For $|x| \leq a$, $f_1(x) - f_o(x) = \varepsilon h_1(x) \geq 0$. Also, the second integral equals zero. Thus $J \geq 0$ and $f_o(x)$ minimizes Fisher's Information over the entire class.

References - Chapter III

1. J. Capon, "On the Asymptotic Efficiency of Locally Optimum Detectors," IRE Trans. Infor. Theo., Vol. IT-7, pp. 67-71, April 1961.

2. R. F. Ingram and R. Houle, "Performance of the Optimum and Several Suboptimum Receivers for Threshold Detection of Known Signals in Additive, White, Non-Gaussian Noise," NUSC Technical Report 6339, Naval Underwater Systems Center, New London, CT, November 1980.

3. P. J. Huber, "Robust Estimation of a Location Parameter," Ann. Math. Stat., Vol. 35, pp. 73-101, March 1964.

4. A. H. El-Sawy and V. D. VandeLinde, "Robust Detection of Known Signals," IEEE Trans. Infor. Theo., Vol. IT-23, pp. 722-727, November 1977.

5. R. J. Marks, G. L. Wise, D. G. Haldeman, J. L. Wited, "Detection in Laplace Noise," IEEE Trans. on Aerospace and Electronic Systems, Vol. AES-14, pp. 866-872, November 1978.

6. H. Cramer, Random Variables and Probability Distributions, Cambridge University Press, Cambridge, 2nd edition, 1962.

7. J. H. Miller and J. B. Thomas, "Numerical Results on the Convergence of Relative Efficiencies," IEEE Trans. on Aerospace and Electron. Systems, Vol. AES-11, pp. 204-209, March 1975.

8. M. Rosenblatt, "A Central Limit Theorem and a Strong Mixing Condition, "Proc. Natl. Acad. Sci., Vol. 42, pp. 43-47, January 1956.

9. E. L. Price and V. D. VandeLinde, "Robust Estimation Using the Robbins-Monro Stochastic Approximation Algorithm," IEEE Trans. Infor. Theo., Vol. IT-25, No. 6, pp. 698-704, November 1979.

10. P. J. Huber, Robust Statistics, John Wiley & Sons, Inc., New York, 1981.

11. J. H. Miller and J. B. Thomas, "Detectors for Discrete-Time Signals in Non-Gaussian Noise," IEEE Trans. Infor. Theo., Vol. IT-18, No. 2, pp. 241-250, March 1972.

12. R. D. Martin and S. C. Schwartz, "Robust Detection of a Known Signal in Nearly Gaussian Noise," IEEE Trans. Infor. Theo., Vol. IT-17, No. 1, January 1971.

13. S. A. Kassam and J. B. Thomas, "Asymptotically Robust Detection of a Known Signal in Contaminated Non-Gaussian Noise," IEEE Trans. Infor. Theo., Vol. IT-22, No. 1, pp. 22-26, January 1976.

14. P. J. Huber, "Fisher Information and Spline Interpolation," The Annals of Statistics, Vol. 2, No. 5, pp. 1029-1033, September 1974.

Chapter IV

AN ADAPTIVE DETECTOR FOR SIGNALS IN NON-GAUSSIAN NOISE

Adaptive detectors are often used when near optimum performance is required for a wide variety of noise environments. They are often more complex than non-adaptive detectors; however, they are inherently capable of handling noise whose statistics are unknown or time-varying. In this chapter an adaptive detector is presented and simulation results are given which indicate that this detector will perform well for several different background noises. Section 1 starts with a description of the detector. Different measures of density function tail behavior are then explored to determine an algorithm which may be used to adapt the detector. In order to test the detector's performance, it is implemented on a digital computer. Section 2 presents the results of this simulation.

Section 1: Development of an Adaptive Detector using the Johnson System as a Noise Model

Figure 4.1 displays the basic structure for an adaptive detector which can be used for the hypothesis testing problem

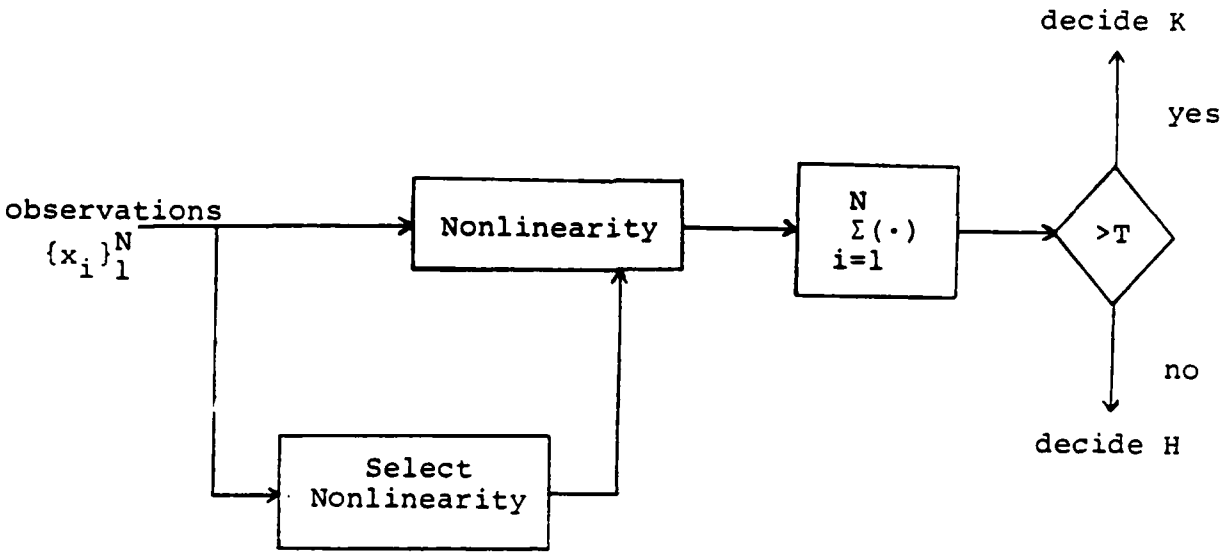


Fig. 4.1 - An Adaptive Detector for Signals in Non-Gaussian Noise

described in Chapter I. Information is extracted from the observations in order to select a nonlinearity to be used for the detection. If one has a reasonable model for the noise, with perhaps a few unknown parameters, the selection of the nonlinearity becomes less difficult. The parameters may be estimated from the observations to determine the density function within the model which most closely approximates the noise. The optimum nonlinearity for this density should be close to optimum for the actual noise.

The Johnson S_u System, presented in Chapter II - Section 3, provides a useful noise model. It was chosen because it is a relatively simple parameterized family of densities which possesses noise blankers for the optimum detector nonlinearities (see Fig. 2.15 and Appendix 2.2). These nonlinearities are nearly linear about the origin. However, large observations are heavily suppressed. The Johnson System for zero mean symmetric distributions [Eq. (2.35)] can be described by two parameters, δ and λ. The shape parameter δ is directly related to the tails of the density. The scale parameter λ can be related to the variance σ^2 by the following equation

$$\lambda = \left[\frac{2\sigma^2}{e^{2/\delta^2} - 1}\right]^{1/2} \tag{4.1}$$

In order to use the Johnson System nonlinearities, an adaptive detector must estimate the parameters δ and λ. Given estimates of δ and σ^2, λ can easily be estimated using Eq. (4.1).

The shape parameter δ can be estimated by utilizing its relationship to tail behavior. Three measures of density function tail behavior will now be described in order to facilitate the selection of the appropriate measure for the estimation of δ.

Measures of Density Function Tail Behavior

It is desirable that a measure of tail behavior be independent of location, be independent of scale, and exist for all distributions. The most common measure of tail behavior, the standardized fourth central moment defined by

$$\beta_2 = \frac{E(x - \mu)^4}{[E(x - \mu)^2]^2} \qquad \text{where } \mu = E(x) \tag{4.2}$$

possesses the first two of these properties. However, any simple function of moments cannot exist for all distributions. This leads to the use of functions of percentage points as a measure of tail behavior. One such measure [1] is

$$\tau = \frac{R(p_1)}{R(p_2)} \tag{4.3}$$

where $R(p) = F^{-1}(p)$, $0 \leq p \leq 1$, and $F(\cdot)$ is the cumulative distribution function. The percentage points p_1 and p_2 are chosen on the tail and on the shoulder of the density, respectively. That is, the point p_1 is selected so that the p_1 quantile, $x_1 = R(p_1)$, increases monotonically

as the tail becomes heavier. The point p_2 is chosen so that the p_2 quantile, $x_2 = R(p_2)$, remains relatively stationary regardless of tail behavior. The measure τ exists for all reasonable distributions and is independent of both location and scale.

Another useful measure of tail behavior [1] is

$$\rho = \frac{R'(p_1)}{R(p_1)} \tag{4.4}$$

where p_1 is chosen on the tail of the density. This ρ possesses many of the same properties as does τ. Its major advantage over τ is that the point on the shoulder of the density, which may be difficult to choose, is not necessary.

The points p_1 and p_2 are chosen on the tail and on the shoulder of the Johnson density system. The inverse of the cumulative distribution function for the Johnson System is

$$x = F^{-1}(p) = \lambda \sinh\left(\frac{\Phi^{-1}(p)}{\delta}\right) \tag{4.5}$$

where $\Phi^{-1}(\cdot)$ is the inverse of the unit Gaussian cumulative distribution function. Figure 4.2 displays x versus p for several values of the parameter δ (λ is selected by Eq. (4.1) so that σ^2 remains constant . For values of $p \geq .995$, as δ decreases so that the tail becomes heavier, the quantile x increases. The larger p is, the easier it becomes to differentiate between the different values of δ. For this

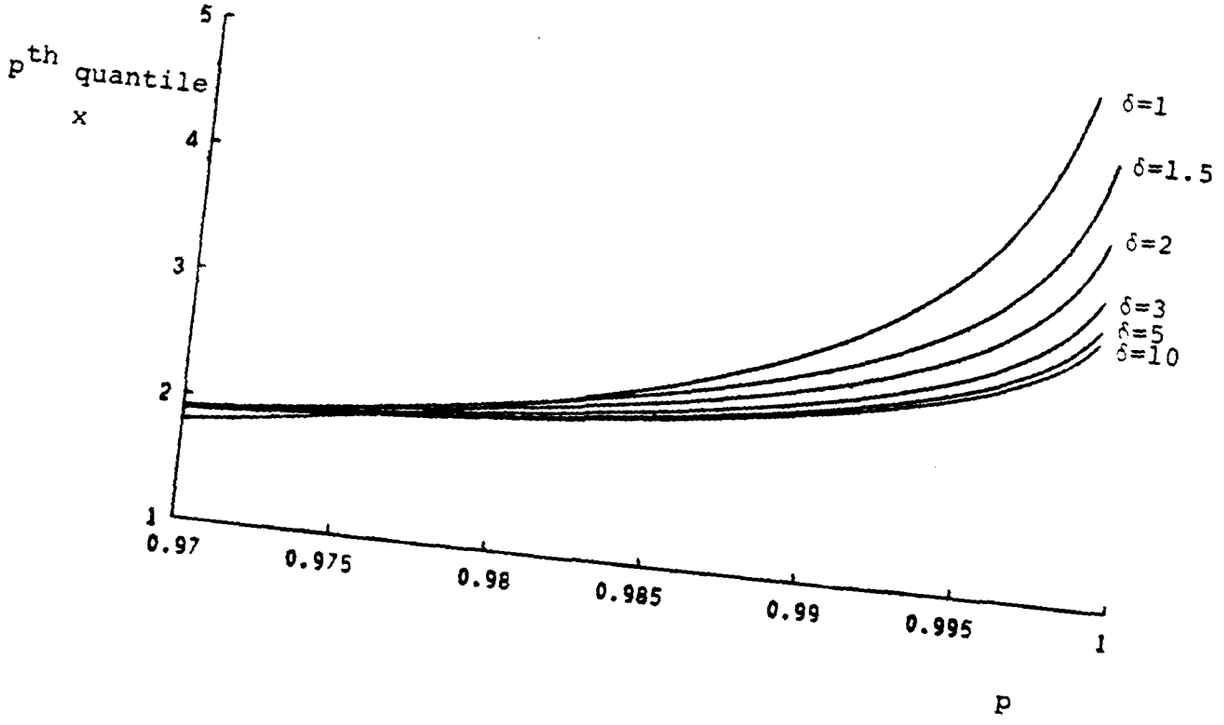


Fig. 4.2 - $x = F^{-1}(p)$ for the Johnson System

reason and to account for detectors operating at a low false alarm rate, the point $p_1 = .9999999$ was chosen. In order to choose p_2 on the shoulder of the system, note that for values of p between approximately .975 and .98 the quantile x is relatively insensitive to the parameter δ. In this study of tail behavior the point $p_2 = .98$ was selected.

The standardized fourth central moment β_2 and the tail measures τ and ρ for the Johnson System, normalized by their values for the Gaussian distribution, are displayed in Fig. 4.3. All three measures increase monotonically as the tails become heavier (δ decreases). Any one of them could be used to estimate a value for δ given a set of observations. Fig. 4.4 shows the relationships between τ and β_2 and between ρ and β_2. Assuming the Johnson System model held, one could measure whichever parameter was easiest and then calculate the remaining.

Due to the importance of Fisher's Information in estimation and detection problems (see Chapter III), the relationships between Fisher's Information and the various tail measures were explored. Fig. 4.5 displays $\sigma^2 I(f)$ versus β_2, τ and ρ (normalized). Given a value for either β_2, τ, or ρ and assuming that the Johnson System noise model is a good approximation to the actual noise environment, one could obtain a unique value for Fisher's Information from Fig. 4.5. This procedure would probably be easier than estimating Fisher's Information directly for an unknown noise distribution.

The usefulness of these three tail measures for a different noise model, the generalized Gaussian noise model presented in

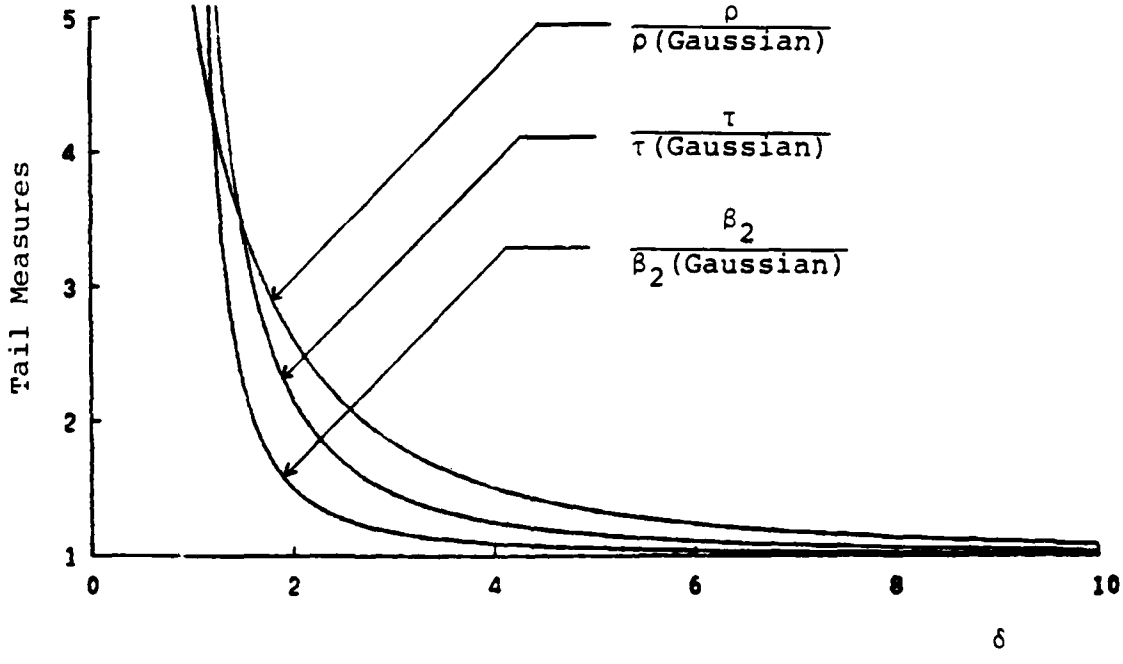


Fig. 4.3 - Johnson's S_u System: β_2, τ, ρ vs. δ

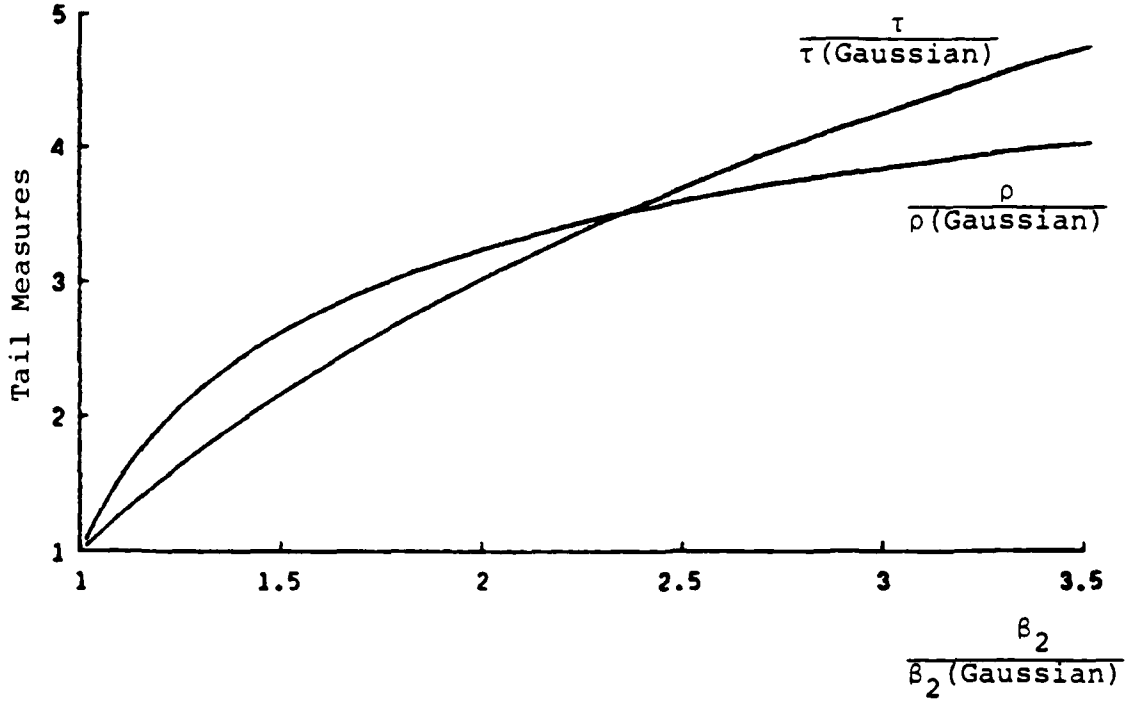


Fig. 4.4 - Johnson's S_u System: τ, ρ vs. β_2

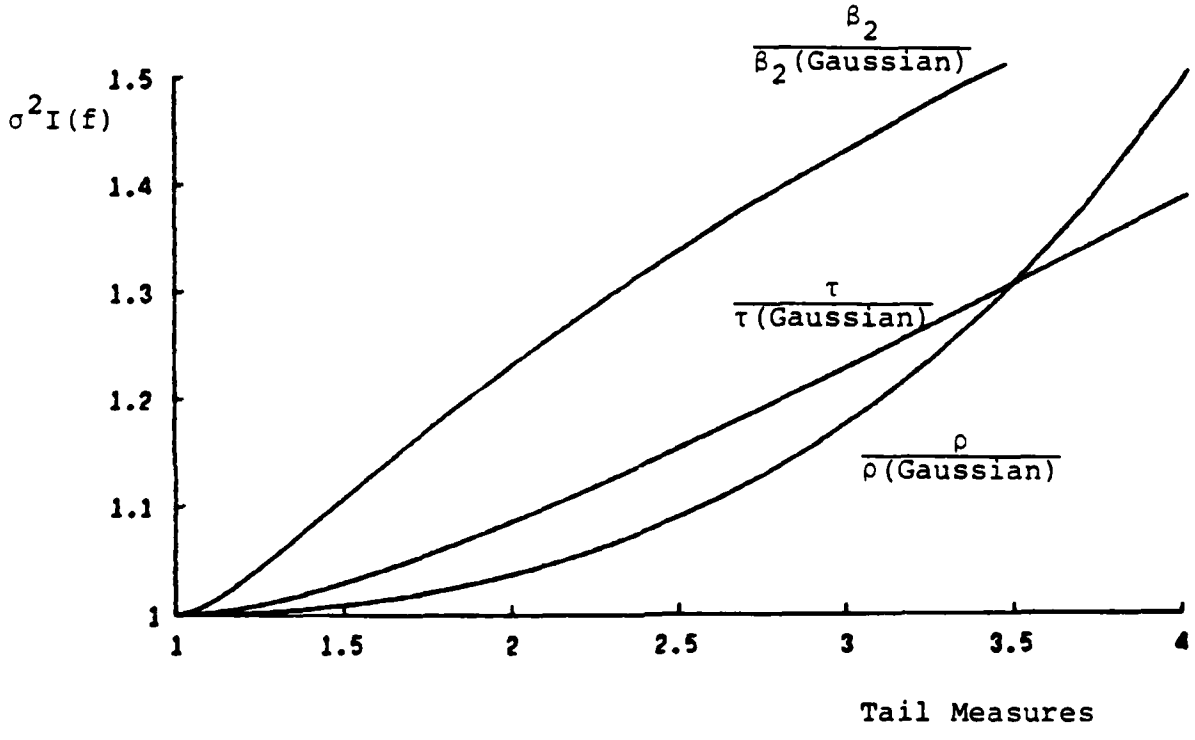


Fig. 4.5 - Johnson's S_u System: Fisher's Information vs. β_2, τ, ρ

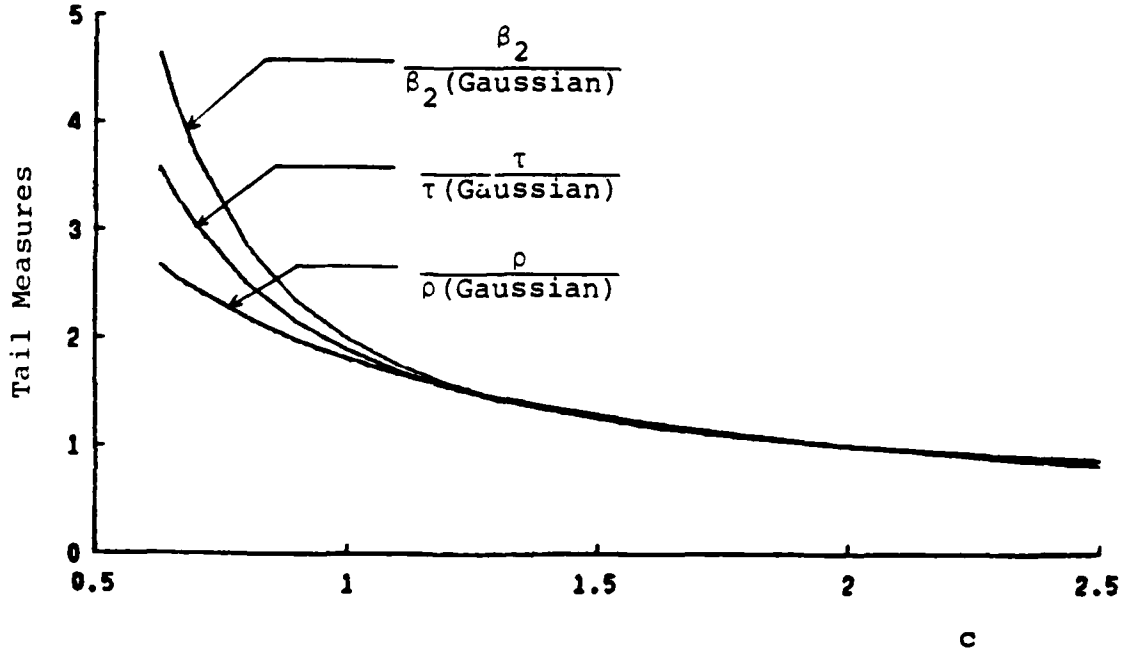


Fig. 4.6 - Generalized Gaussian Noise: β_2, τ, ρ vs. c

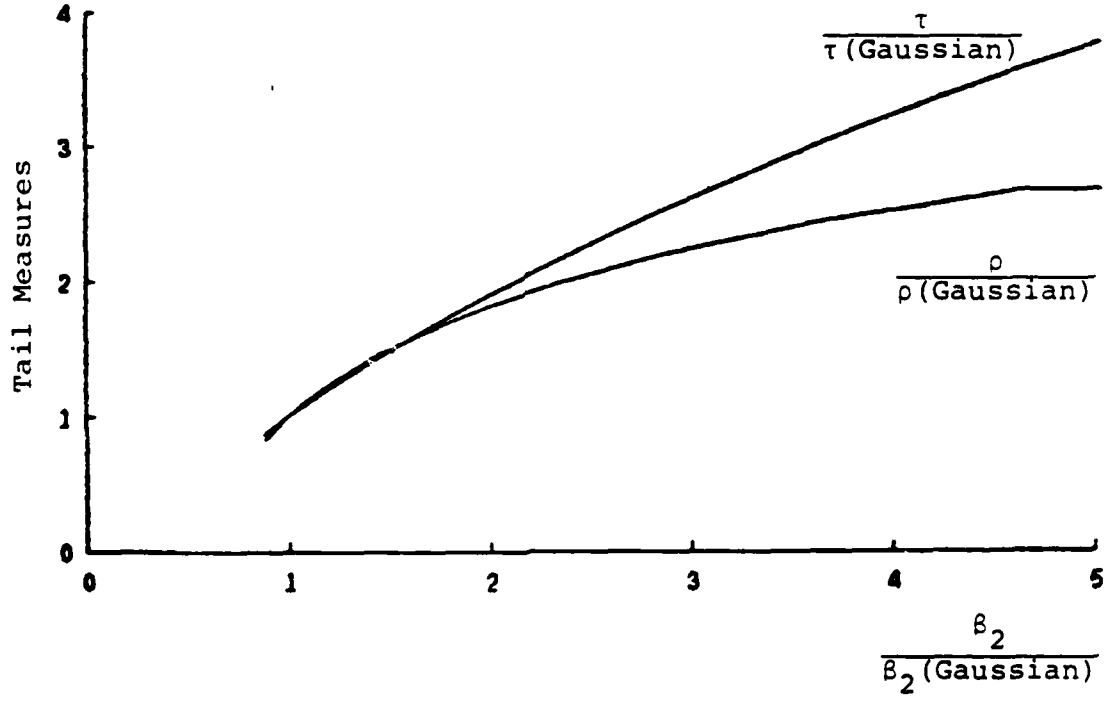


Fig. 4.7 - Generalized Gaussian Noise: τ, ρ, vs. β_2

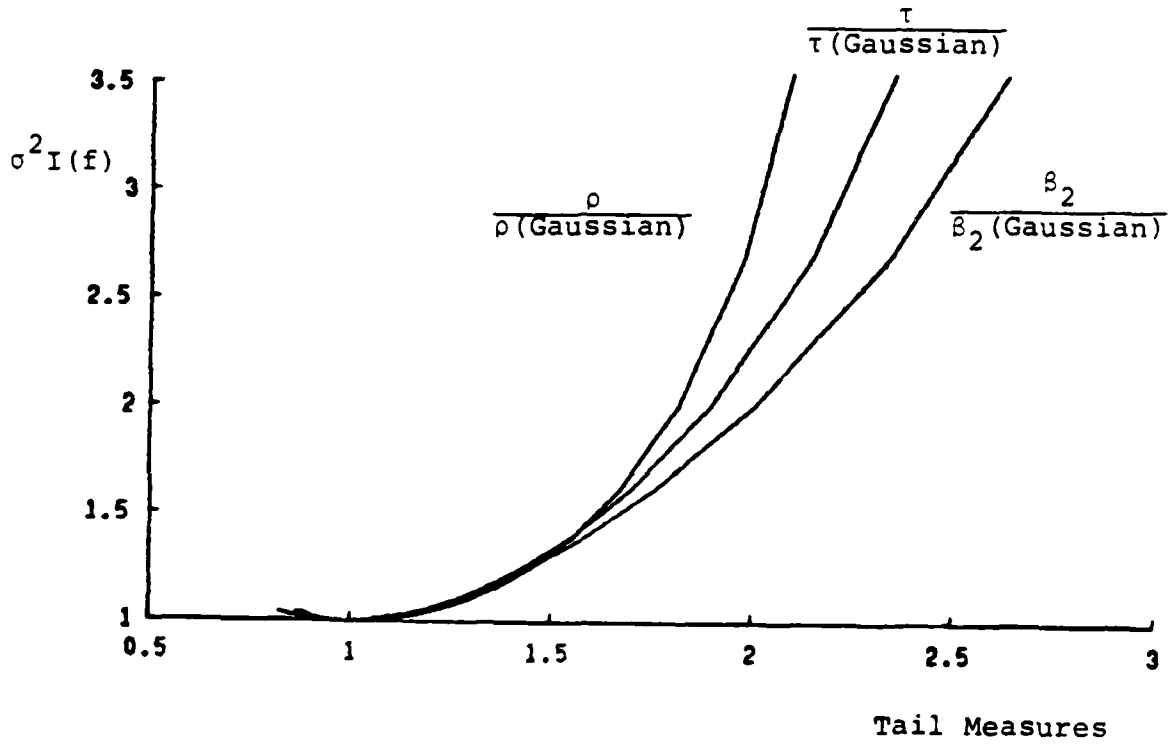


Fig. 4.8 - Generalized Gaussian Noise:
Fisher's Information vs. β_2, τ, ρ

Chapter II - Section 3, was also considered. Fig. 4.6 shows β_2, τ and ρ versus the density system parameter c. As for the Johnson System, the three measures increase monotonically as the tail becomes heavier (c decreases). Through these simple relationships a value for c could easily be estimated by measuring either β_2, τ or ρ. Fig. 4.7 displays the monotone relationships which enable one to estimate β_2, τ and ρ from each other. A plot of Fisher's Information versus β_2, τ and ρ is given in Fig. 4.8.

An additional use for these tail measures, which was only briefly explored, would be as tools for the comparison of noise models. Various densities, whose tail behavior are identical as measured by either β_2, τ or ρ, could be compared in terms of, for example, Fisher's Information or efficacy with various detectors.

Implementation of the Johnson System Adaptive Detector

Any one of the three measures of density function tail behavior could be used to estimate the Johnson System shape parameter δ. The measure τ was selected because it appeared to be the easiest to estimate from data. One need only measure two quantiles and take the ratio. The measure β_2 would require taking a fourth moment. This would probably be more difficult in most applications. Although ρ requires only one quantile, the rate of increase in the quantile would also have to be calculated. This could be achieved by measuring a few points above and below the quantile. This method, however, may not be very accurate.

The point $p_1 = .9999999$, selected previously for the study of the tail measures, can no longer be used. This quantile is too far out on the tail to be estimated accurately with reasonable sample sizes. Checking Fig. 4.2 one finds that the smallest value of p which will still allow one to differentiate accurately between the various values of δ is .995. Consequently, $p_1 = .995$ was chosen. The point p_2 was changed from .98 to .975 so that the two points, p_1 and p_2, would remain an adequate distance apart. Choosing $p_1 = .995$ and $p_2 = .975$, one can write τ for the Johnson System as

$$\tau = \frac{\sinh(2.576/\delta)}{\sinh(1.960/\delta)} \tag{4.6}$$

This equation is plotted in Fig. 4.9. The curve is similar to the one in Fig. 4.3.

To estimate δ given a set of observations, τ is first estimated from the two quantiles

$$\tau = \frac{\text{the 99.5\% point}}{\text{the 97.5\% point}} \tag{4.7}$$

and then δ is estimated using Eq. (4.6). To estimate the scale parameter λ, the sample variance is first formed

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2 \tag{4.8}$$

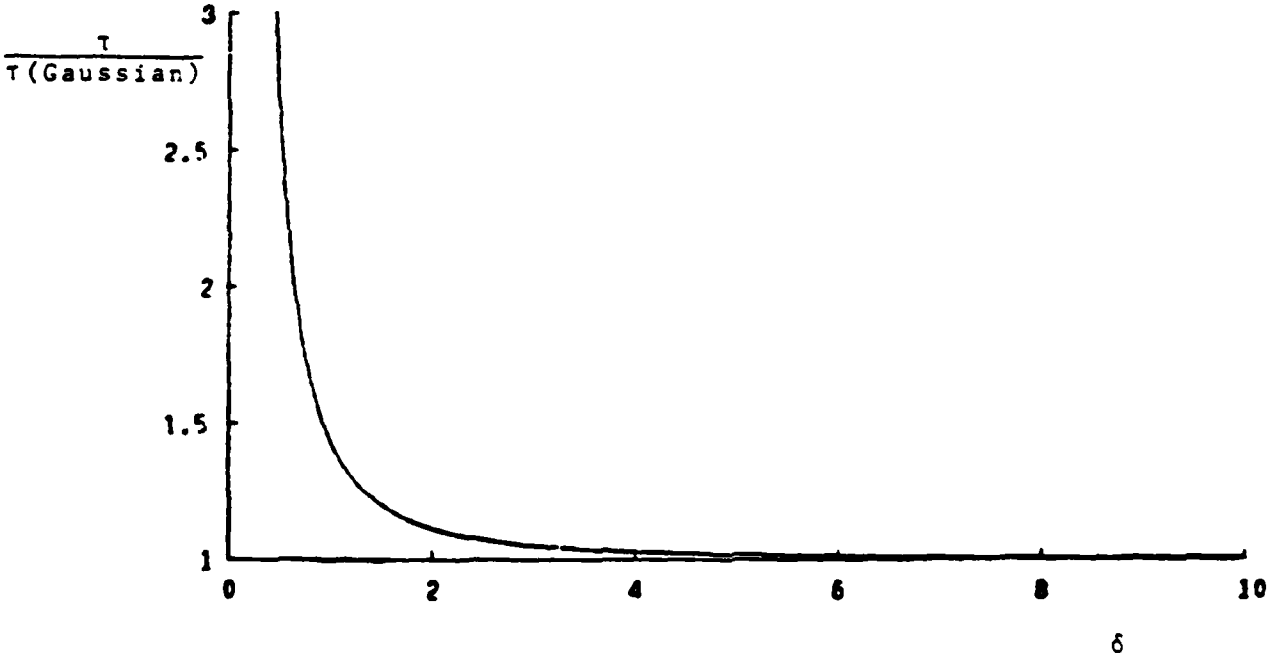


Fig. 4.9 - τ vs. δ for the Johnson System

where $\bar{x}$ is the sample mean. Then, using $\hat{\sigma}^2$ and the estimate for δ, an estimate for λ may be found from Eq. (4.1). Using these estimates for δ and λ the Johnson System density which most closely approximates the noise can be found. The optimum detector for this density should be close to optimum for the actual noise.

Fig. 4.10 displays the block diagram for an implementation of a Johnson System adaptive detector. The detector is of the same form as that shown in Fig. 4.1. Twenty samples were used to make each decision. After every 10,000 samples, τ and the noise variance are estimated and new estimates of δ and λ are formed. The appropriate nonlinearity from the Johnson System is then selected for use on the next 10,000 samples. Simulation studies have been performed with a variety of noise densities on the input. The results from these studies are presented in the next section.

A key problem with the adaptive detector of Fig. 4.10 is that as the input changes and the detector adapts, the probability of false alarm will change unless the threshold is adjusted properly. To eliminate this difficulty, a noisy reference channel was added (Fig. 4.11). The reference channel is a stream of samples with the same distribution function as the noise on the actual channel. However, the two channels are statistically independent. A detector, identical to the one on the actual channel, is placed on the reference channel. From its output, an estimate for the probability of false alarm ($\hat{\alpha}$) can be found. The threshold can then be adjusted, in both channels, to attempt to achieve $\hat{\alpha} = 0.1$. Also, an estimate

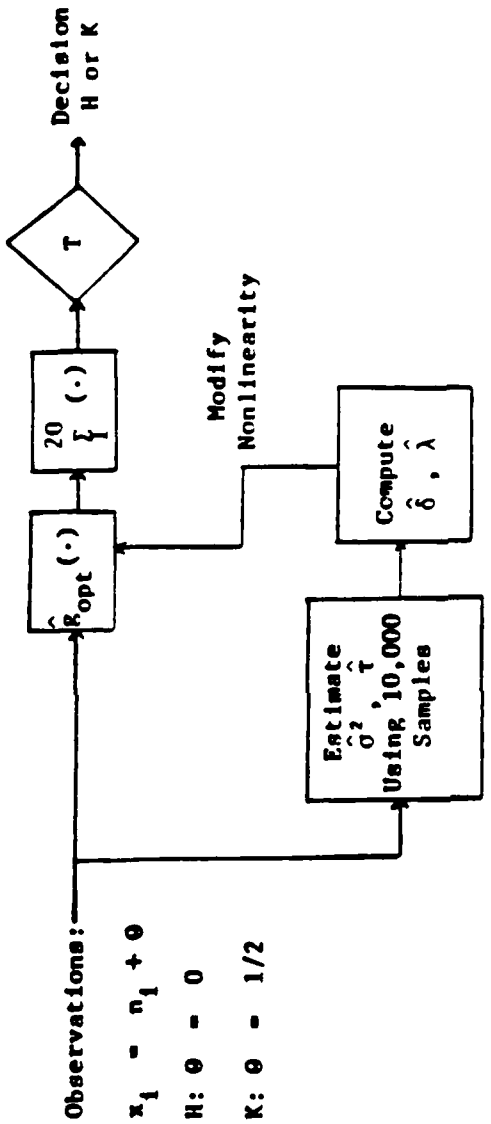


Fig. 4.10 - An Adaptive Detector Based on the Johnson System

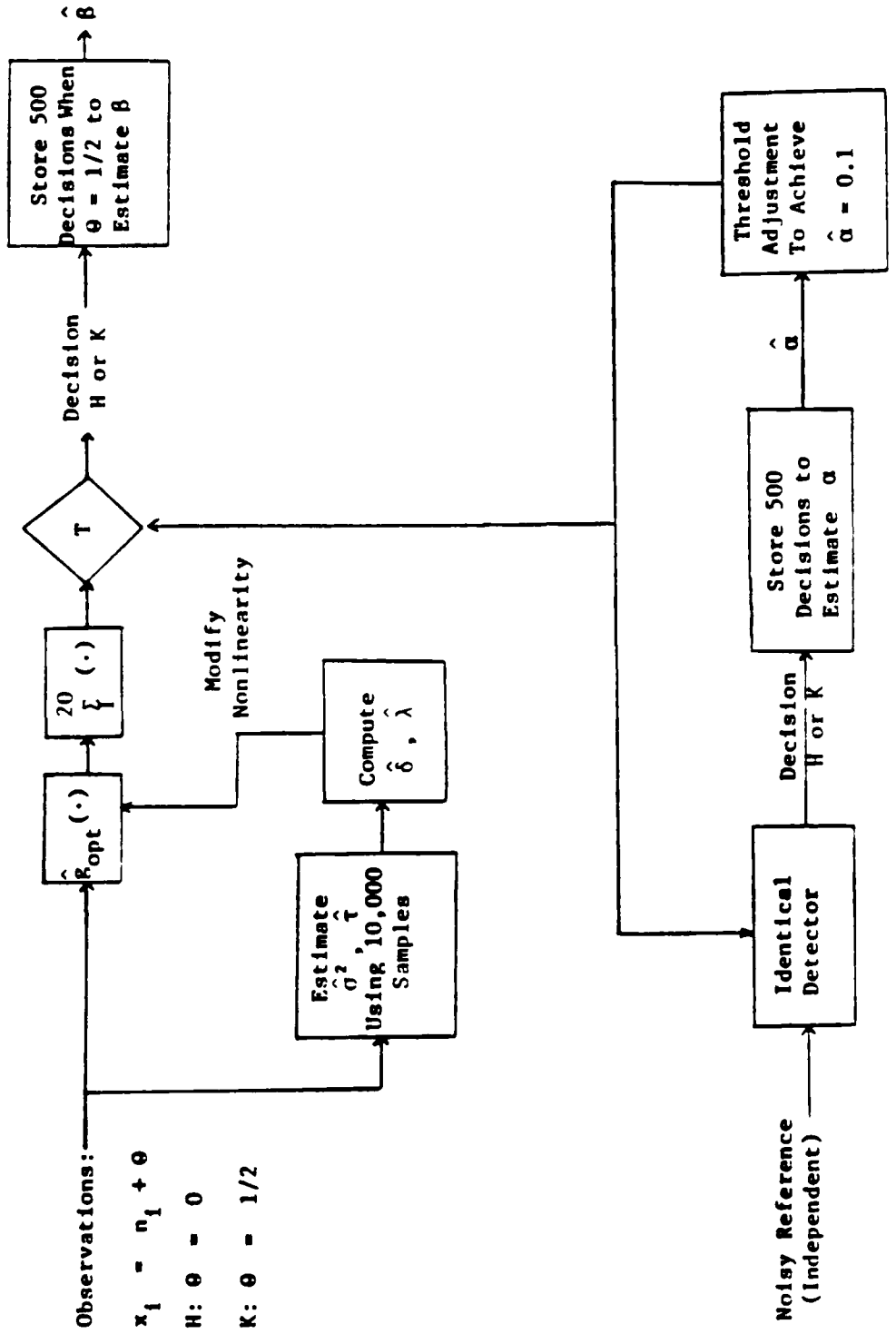


Fig. 4.11 - An Adaptive Detector Based on the Johnson System, Noisy Reference Channel Added.

of the probability of detection ($\hat{\beta}$) can be obtained by counting the number of correct decisions made when the signal is present.

Section 2: Simulation Results

The linear detector, various nonlinear optimum detectors, and the adaptive detector which was described in Section 1 have all been simulated on a digital computer. Various noise backgrounds were considered. All detectors utilized a noisy reference channel to aid in the adjustment of the threshold. The detectors attempted to maintain a probability of false alarm (α) equal to 0.1. The following graphs display estimates of the probability of false alarm (α), the probability of detection (β), and the total probability of error (P_E) for the various detectors and noise situations. The threshold levels (T) are also reported. Thirty estimates of these values are displayed. Each estimate represents 500 decisions made by the detector, 20 samples per decision. The first few estimates were taken before the threshold had stabilized. For this reason, when averaging these numbers the first ten estimates should be ignored.

Figs. 4.12 through 4.15 display α, β, P_E and T for both the optimum and adaptive detectors operating in Gaussian noise. In this case the optimum detector is just the linear detector. Due to the statistical fluctuations of the input, it

is hard to determine precisely the performance of these detectors. However, it appears that their performances are nearly the same.

Figs. 4.16 through 4.19 display α, β, P_E, and T for the optimum, linear, and adaptive detectors in Laplace or double exponential noise. The Laplace distribution has heavier tails than the Gaussian. The optimum detector is the amplifier limiter. The adaptive detector performs nearly as well as the optimum detector (see Fig. 4.17 or 4.18). The linear detector is about 10% worse.

When Johnson noise is on the input of the adaptive detector, one would expect it to perform especially well. When it properly estimates the parameters δ and λ, it will be using the actual optimum detector. Figs. 4.20 through 4.23 show that this did occur. The adaptive detector's performance is nearly identical to that of the optimum detector. Again note that the linear detector is a rather poor detector for this heavy tailed density.

A Gaussian-Gaussian mixture (presented in Chapter II - Section 3) was used to achieve a density with still heavier tails. One Gaussian distribution, the contaminant, was given a variance one hundred times larger than that of the background Gaussian distribution. The contaminant occurred with probability 0.1. The resulting density was normalized to have variance unity. Figs. 4.24 through 4.26 show the performances of the optimum, linear and adaptive detectors for this heavy tailed density. Fig. 4.27 displays the thresholds. One can easily see that the linear detector does not perform very well.

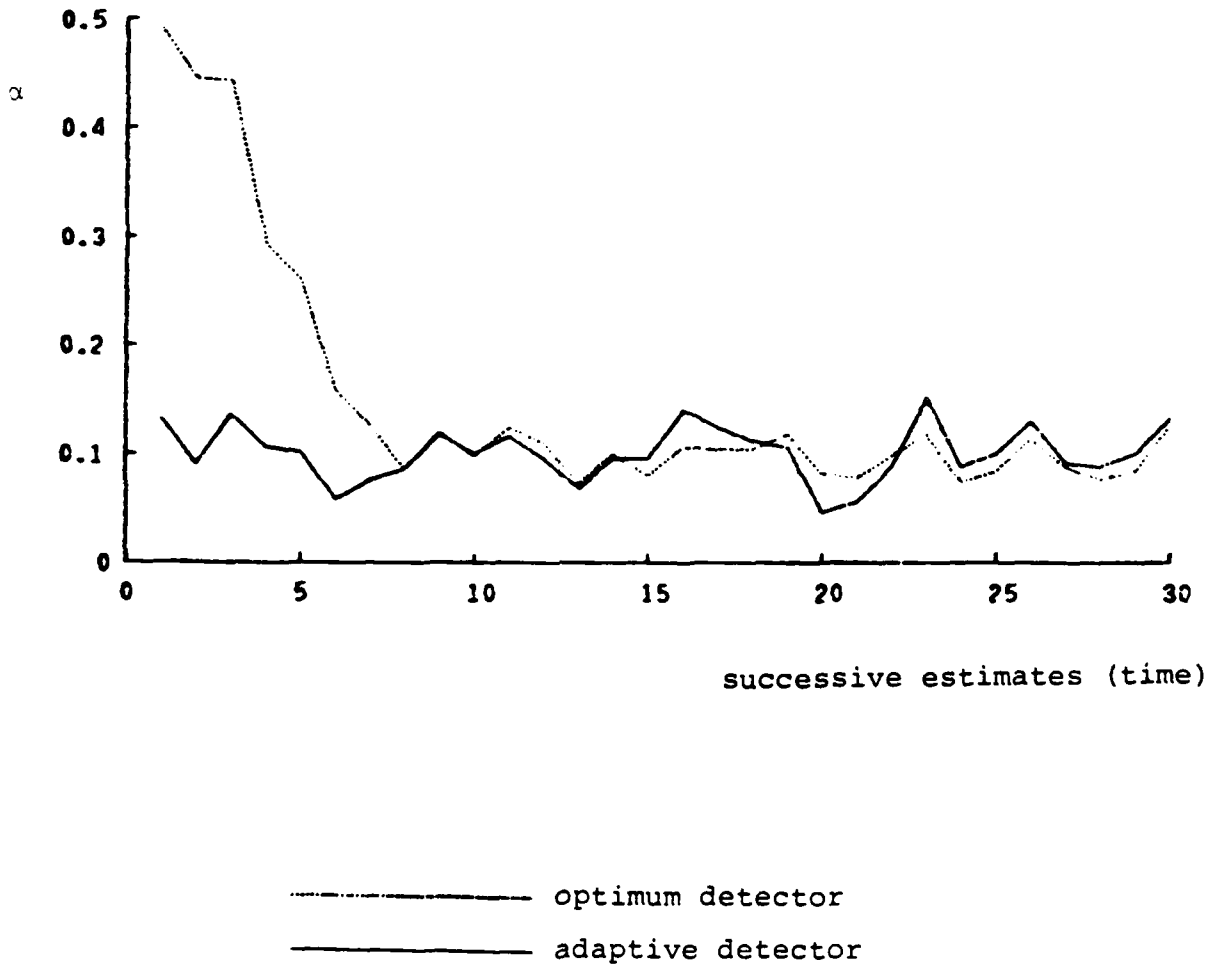


Fig. 4.12 - Probability of False Alarm: Gaussian Noise Variance 1, Signal Strength = 0.5

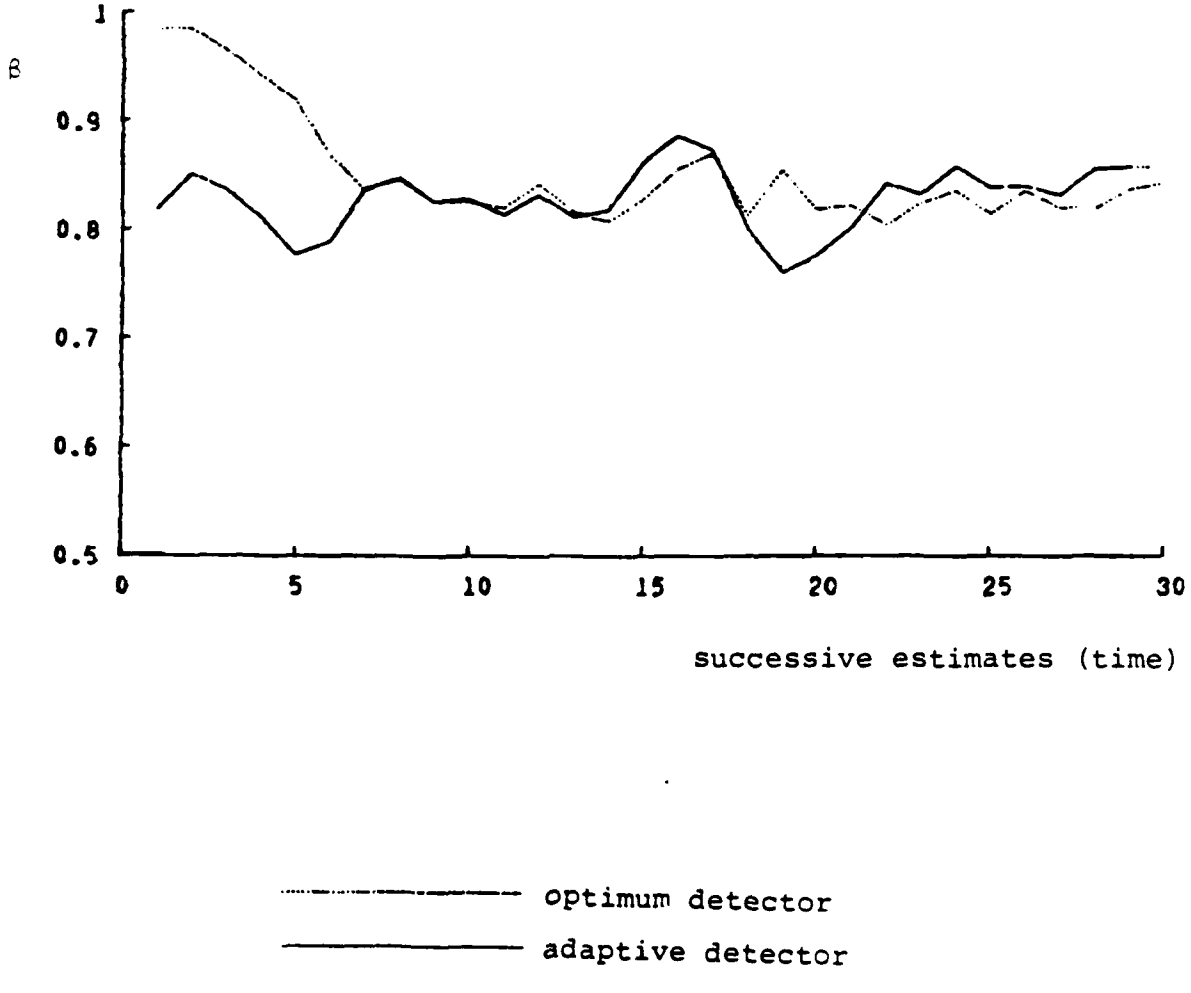


Fig. 4.13 - Probability of Detection: Gaussian Noise
Variance 1, Signal Strength = 0.5

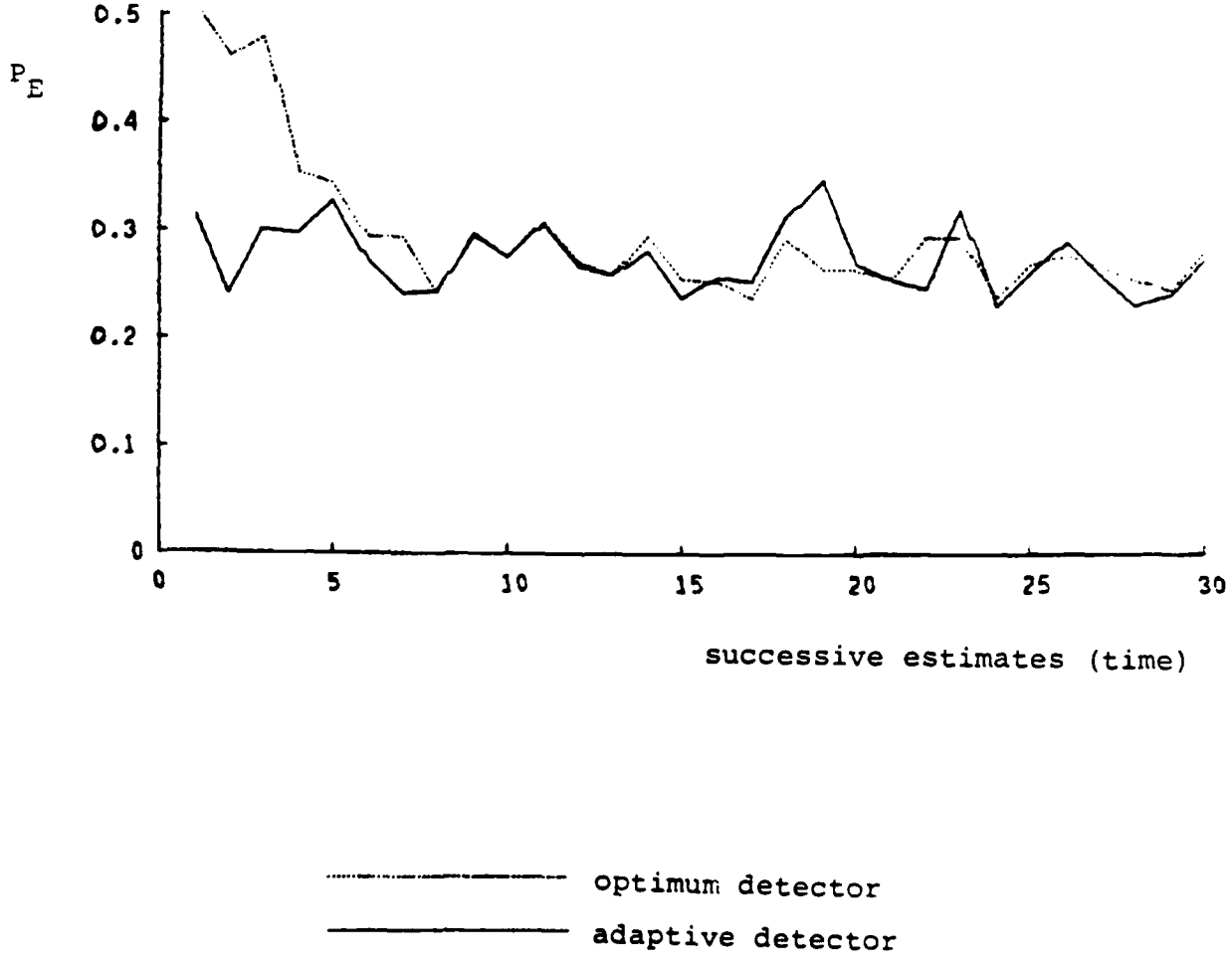


Fig. 4.14 - Total Probability of Error: Gaussian Noise Variance 1, Signal Strength = 0.5

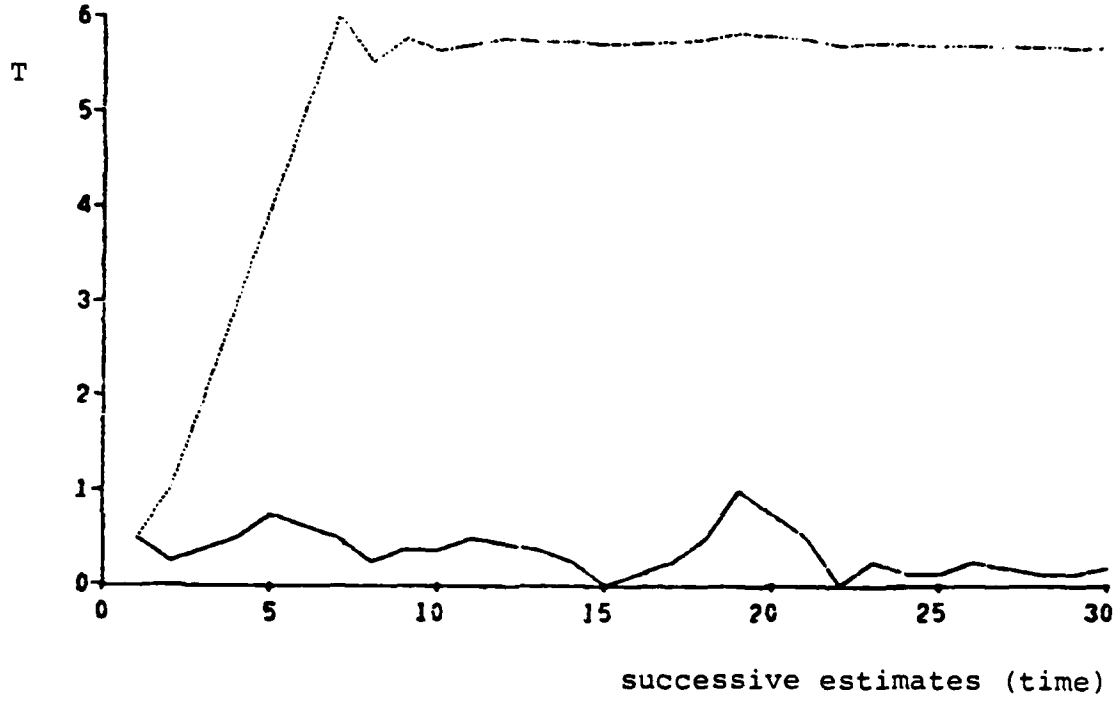


Fig. 4.15 - Threshold Levels: Gaussian Noise
Variance 1, Signal Strength = 0.5

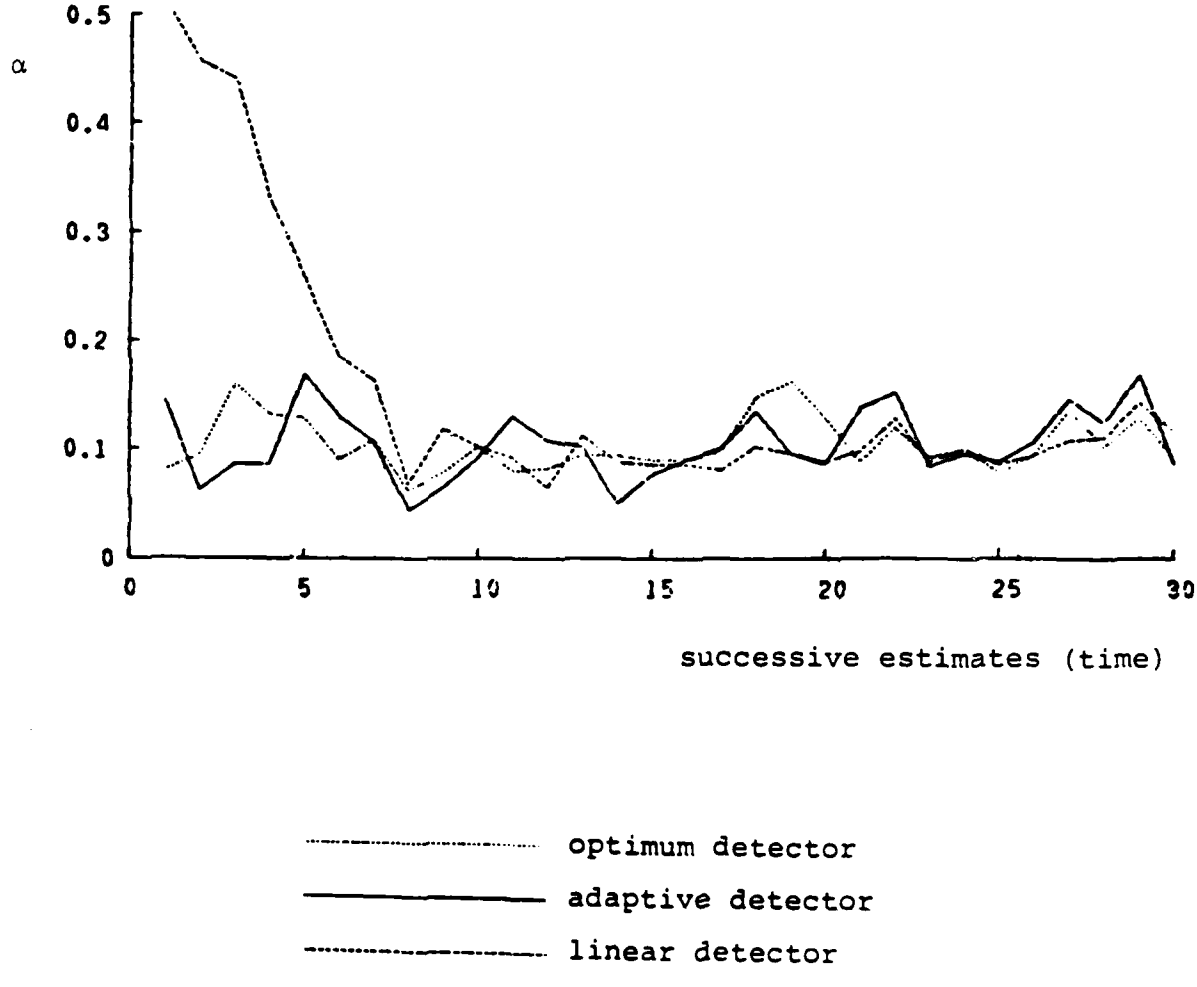


Fig. 4.16 - Probability of False Alarm: Laplace Noise Variance 1, Signal Strength = 0.5

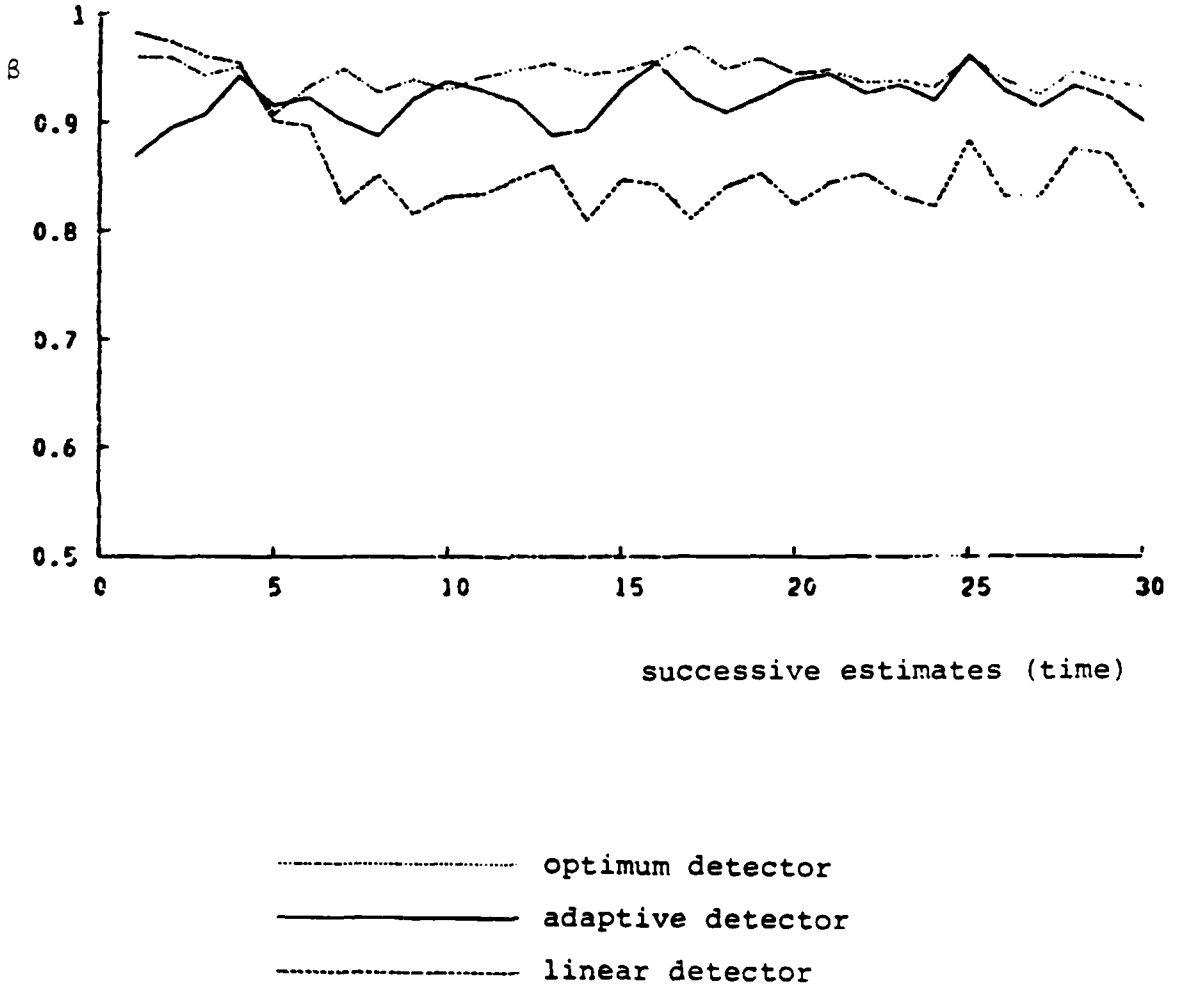


Fig. 4.17 - Probability of Detection: Laplace Noise Variance 1, Signal Strength = 0.5

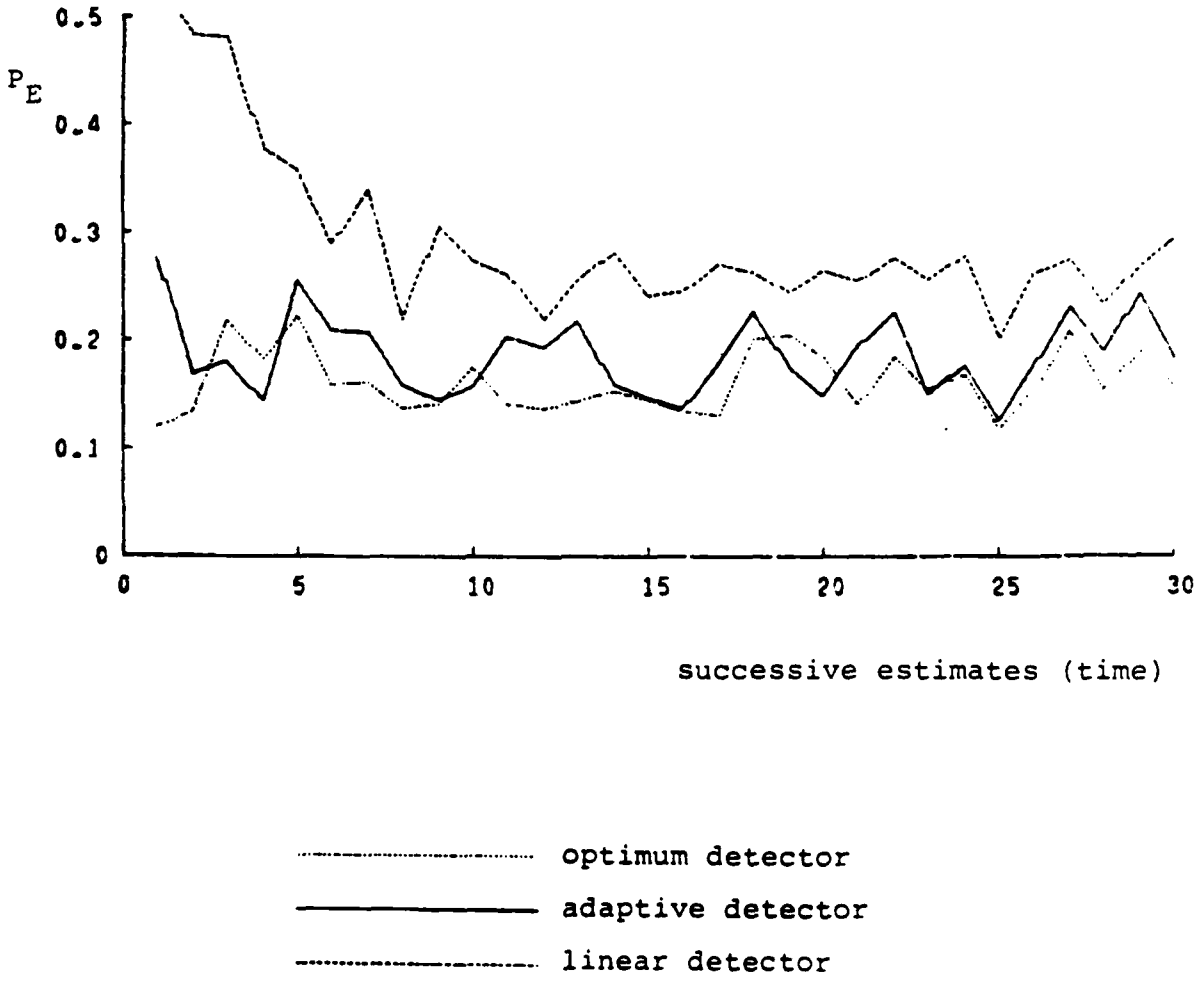


Fig. 4.18 - Total Probability of Error: Laplace Noise Variance 1, Signal Strength = 0.5

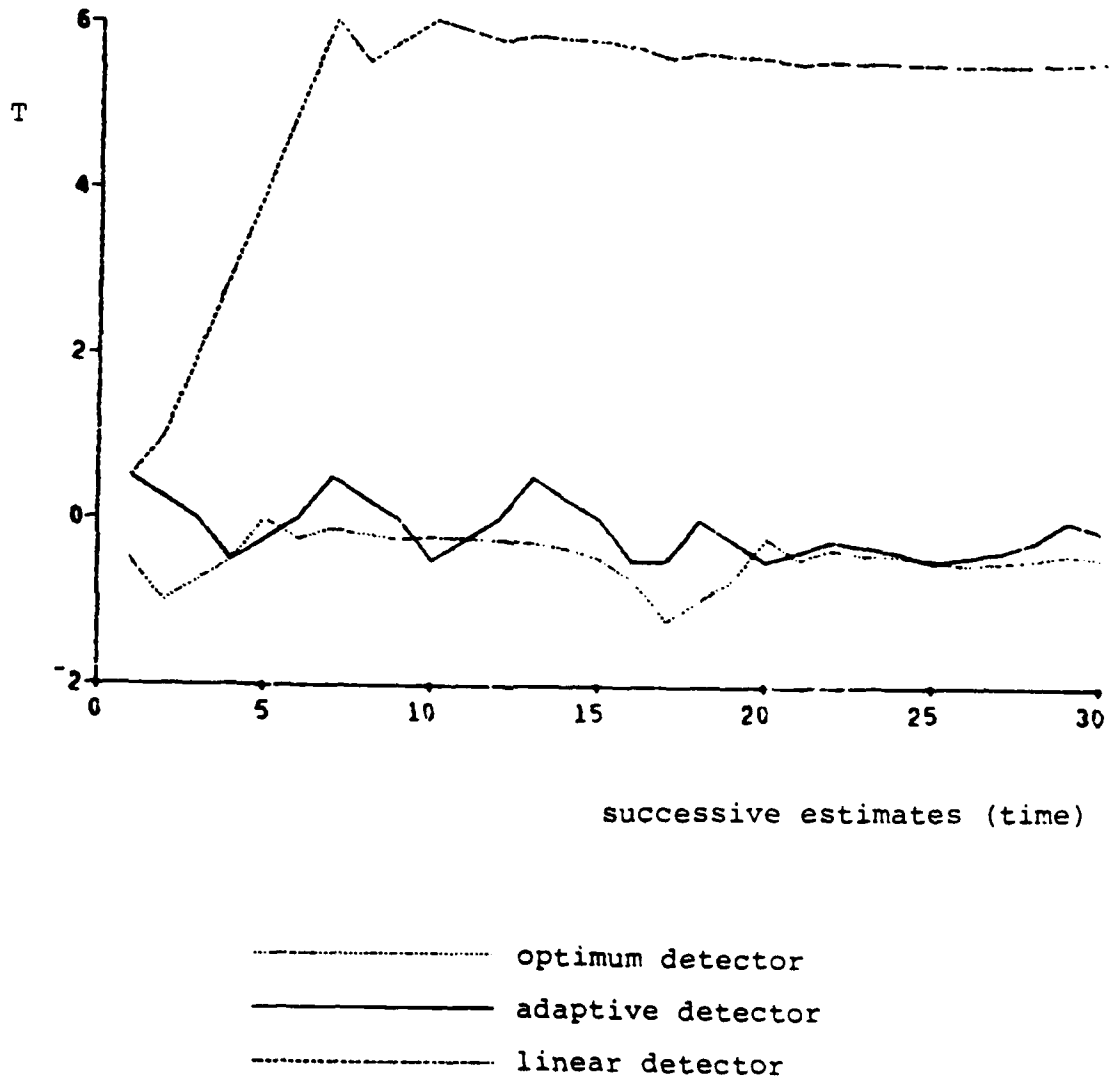


Fig. 4.19 - Threshold Levels: Laplace Noise
Variance 1, Signal Strength = 0.5

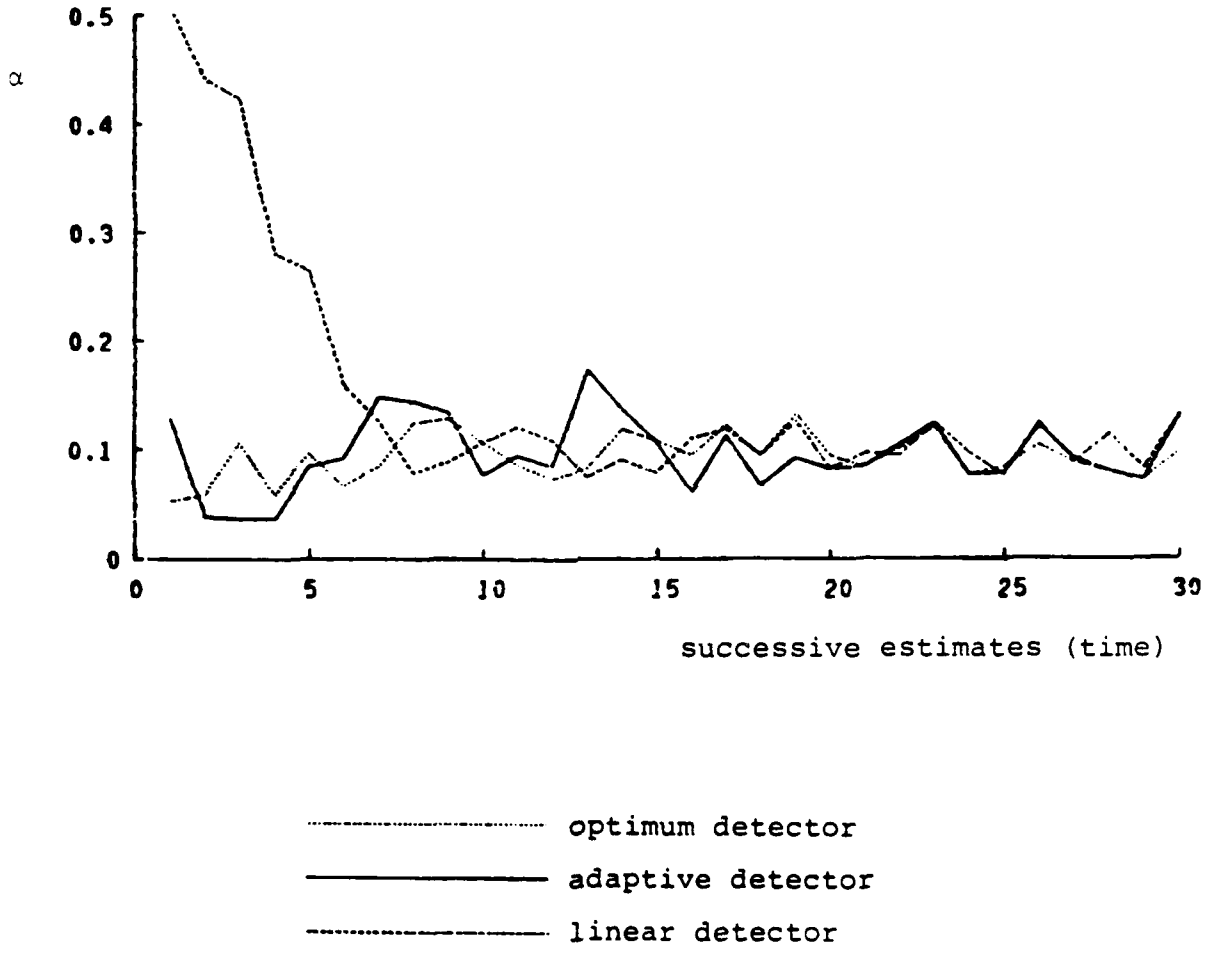


Fig. 4.20 - Probability of False Alarm: Johnson Noise
δ = 1, Variance 1, Signal Strength = 0.5

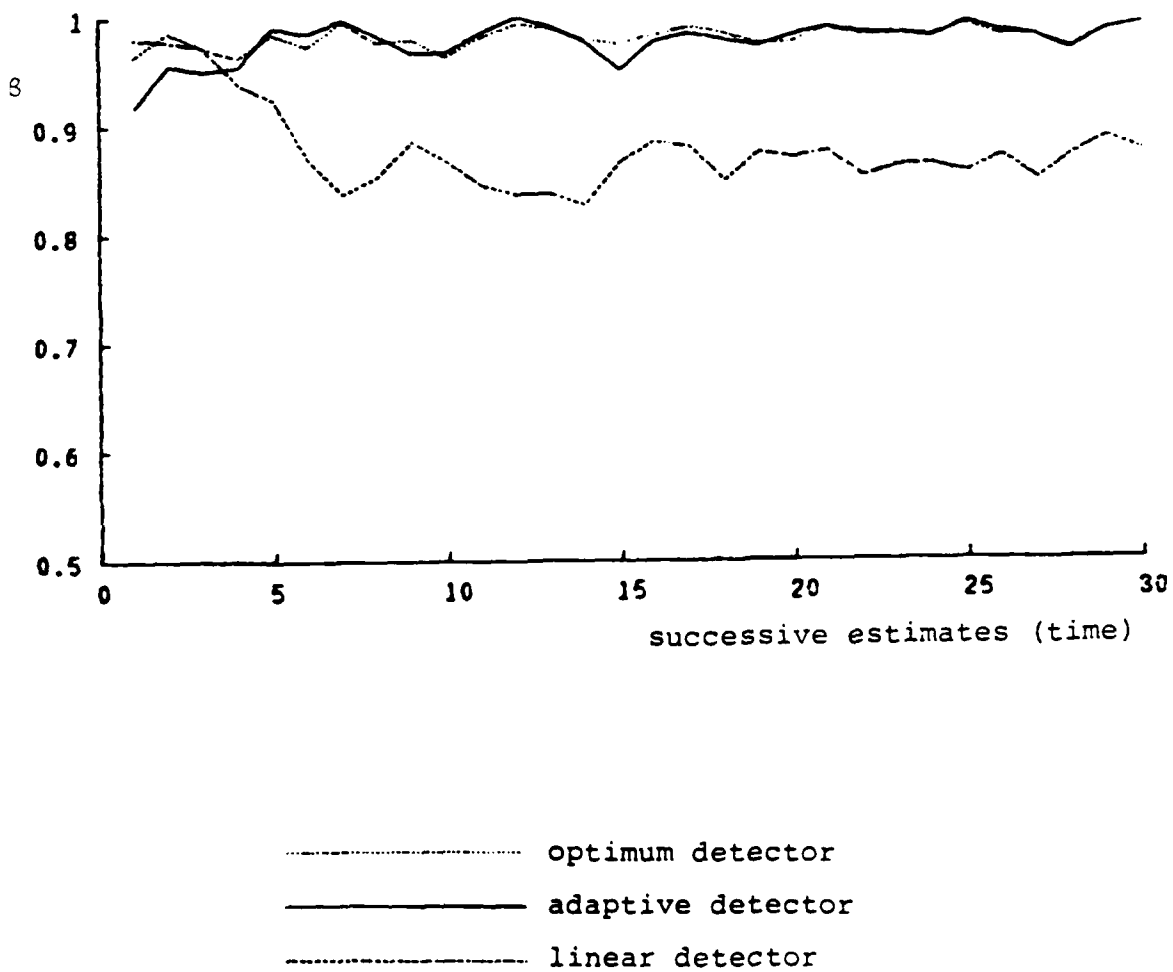


Fig. 4.21 - Probability of Detection: Johnson Noise
δ = 1, Variance 1, Signal Strength = 0.5

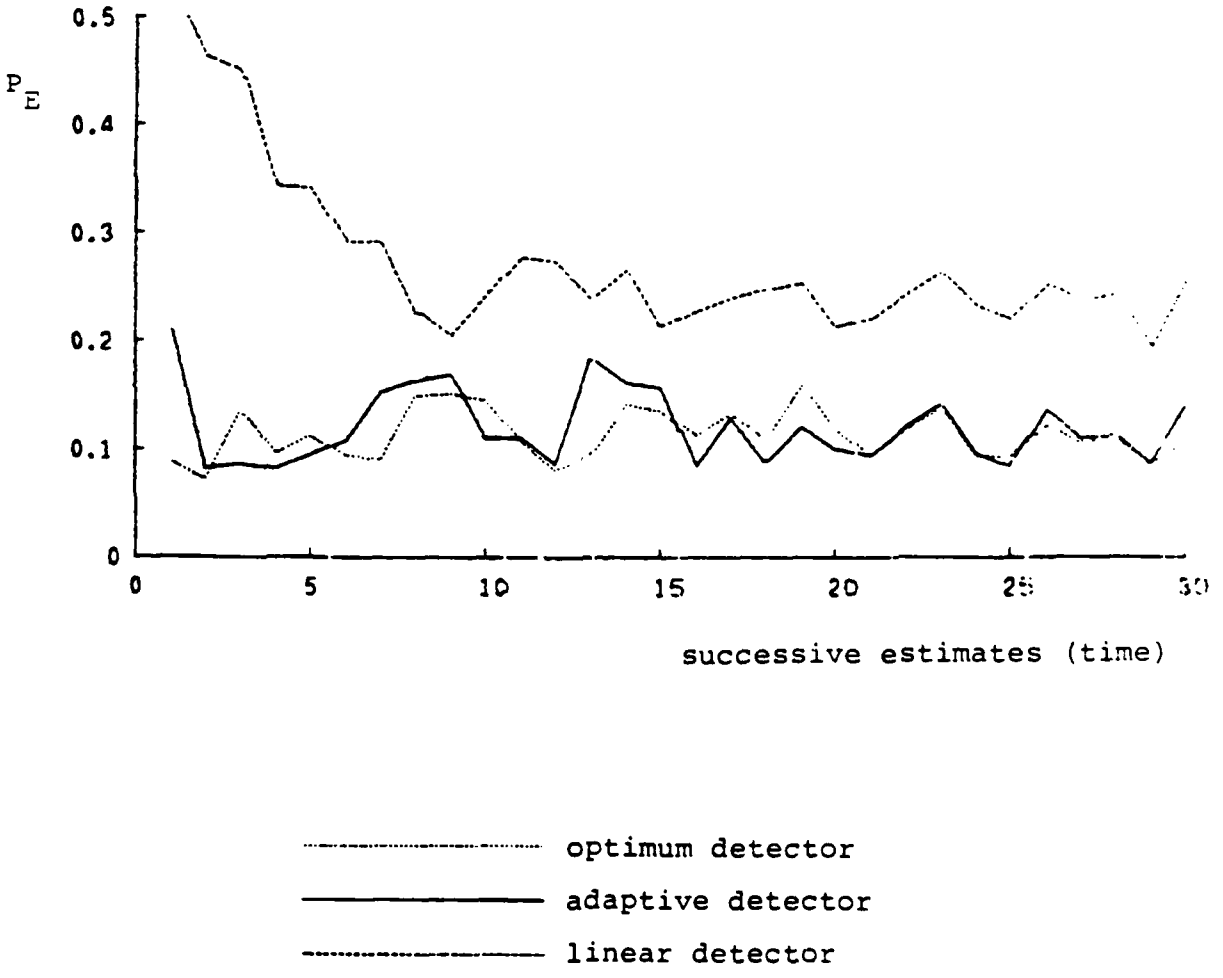


Fig. 4.22 - Total Probability of Error: Johnson Noise
δ = 1, Variance 1, Signal Strength = 0.5

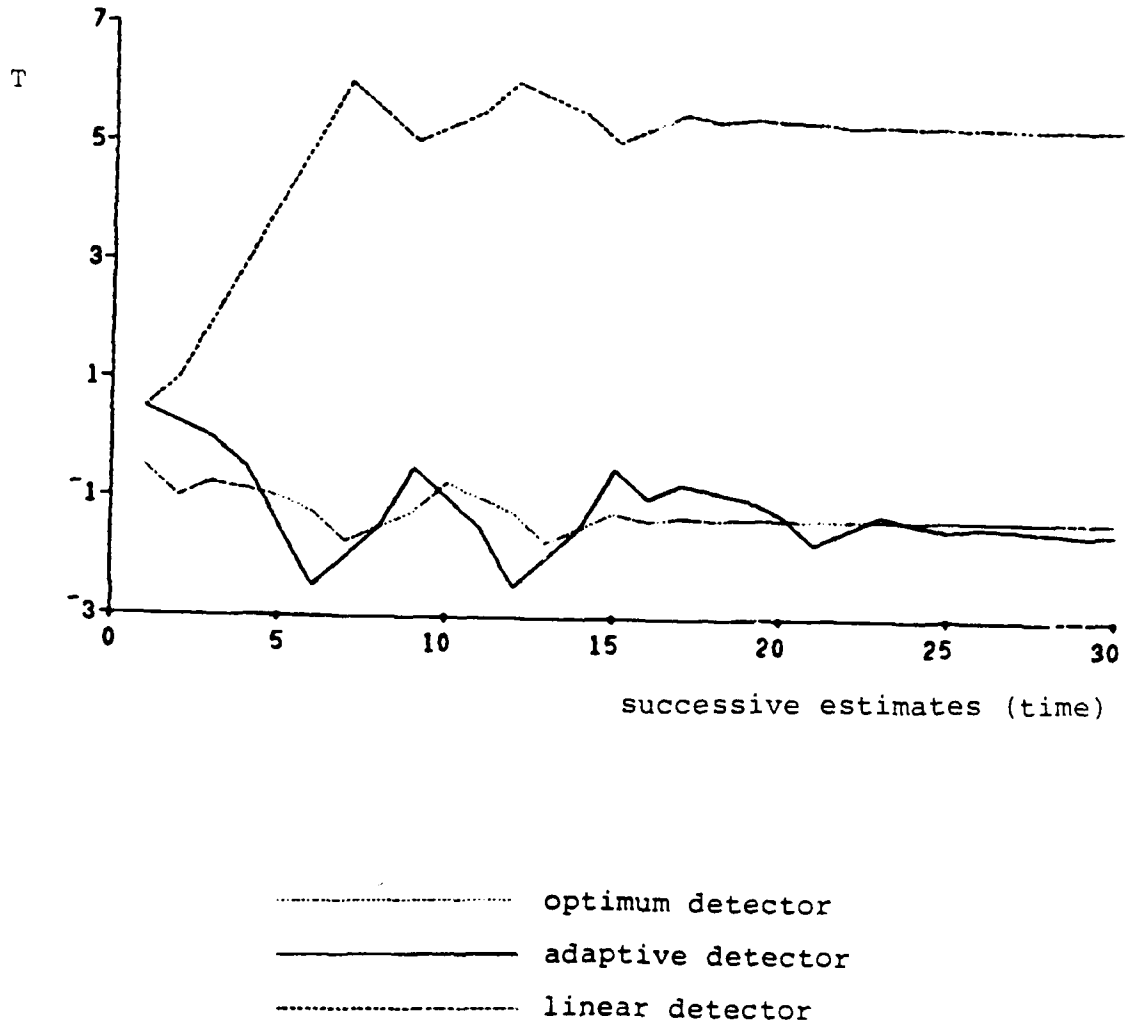


Fig. 4.23 - Threshold Levels: Johnson Noise
δ = 1, Variance 1, Signal Strength = 0.5

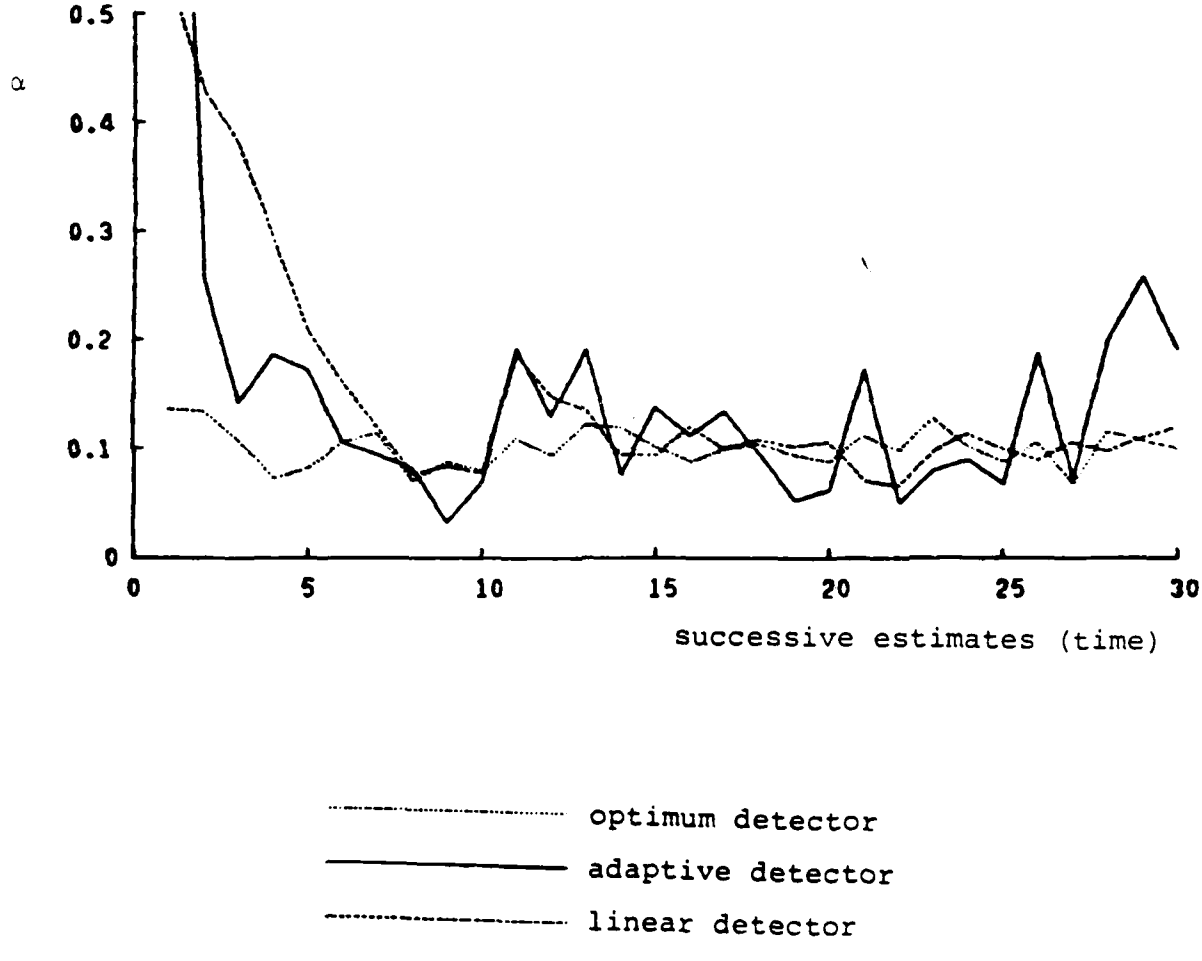


Fig. 4.24 - Probability of False Alarm: Gaussian-Gaussian Mixture Variance 1, Signal Strength = 0.5

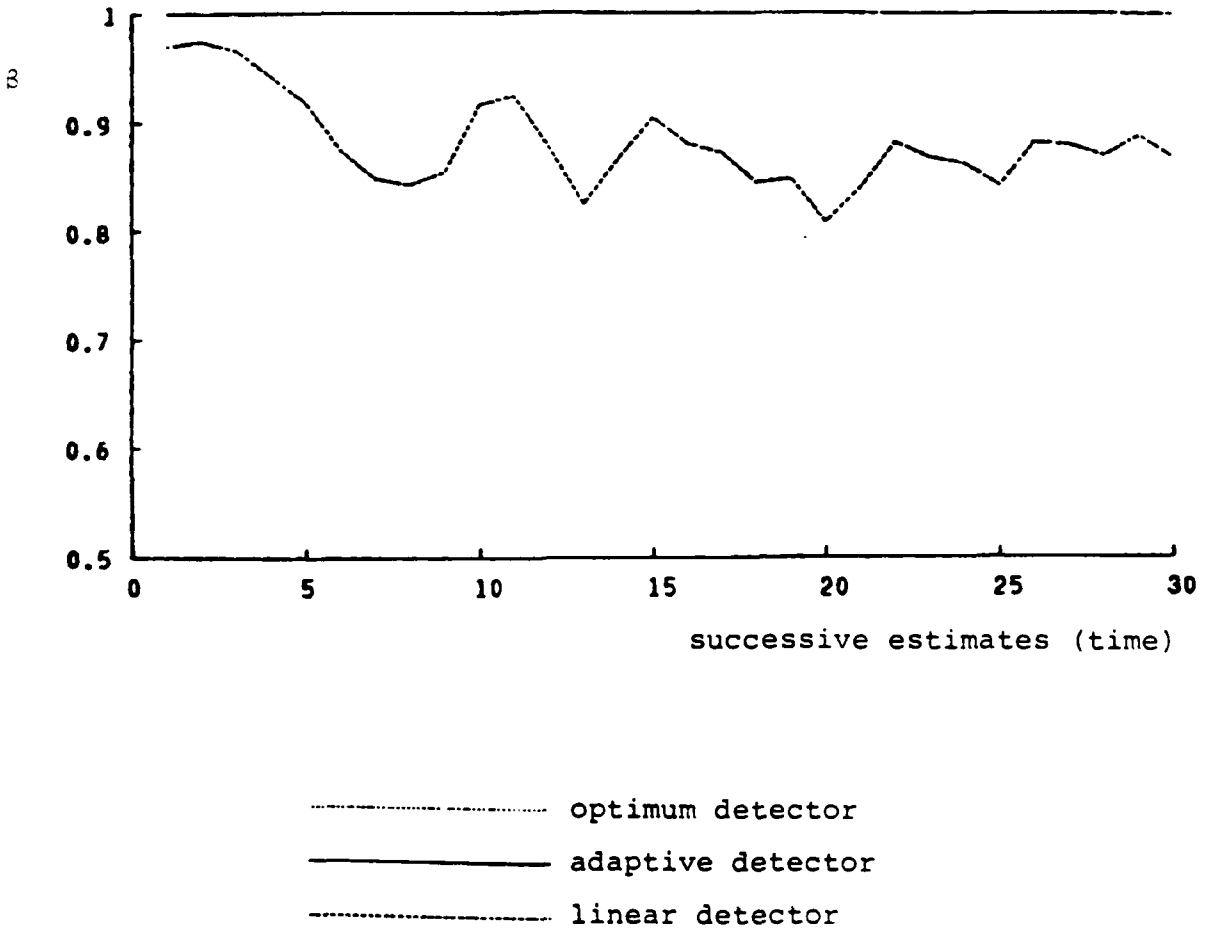


Fig. 4.25 - Probability of Detection: Gaussian-Gaussian Mixture
Variance 1, Signal Strength = 0.5

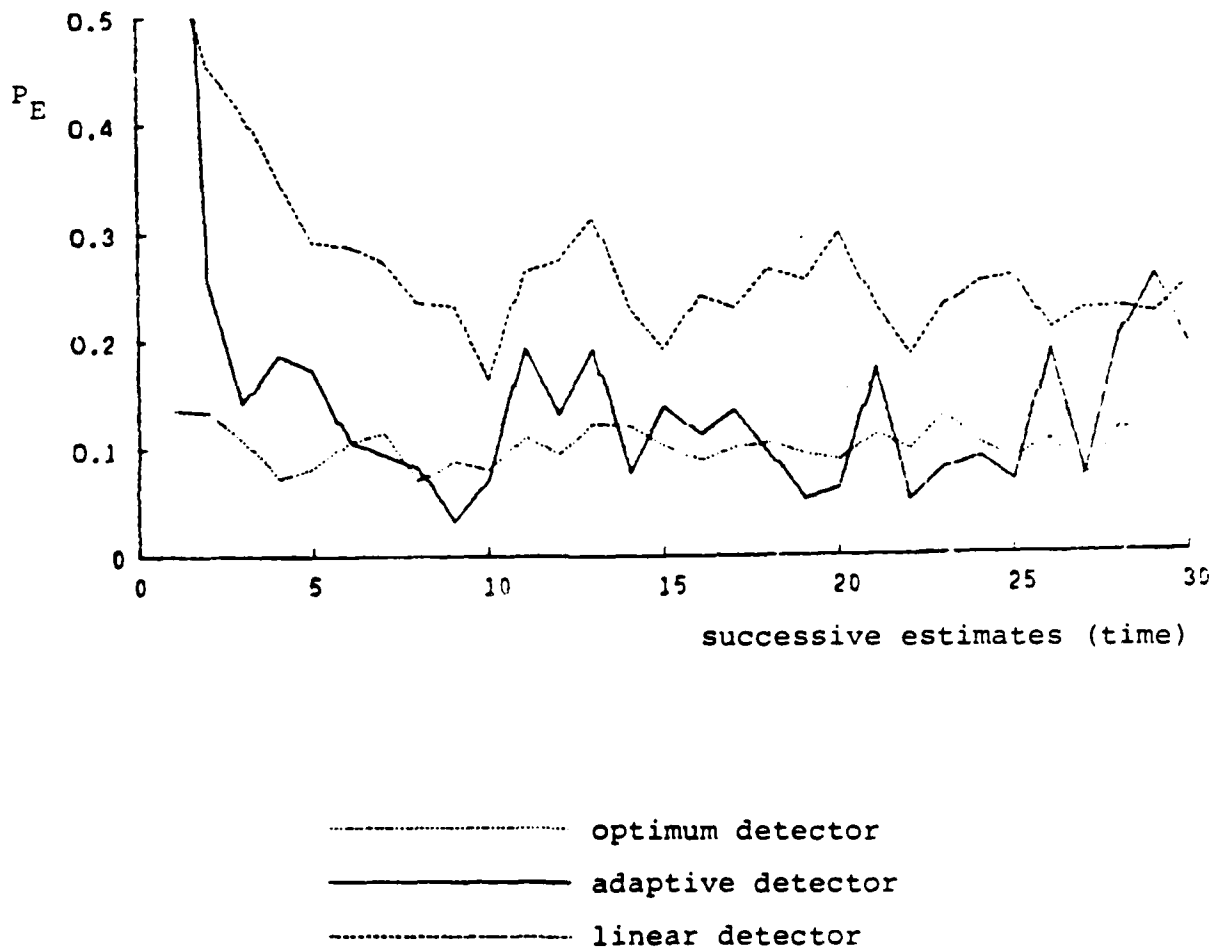


Fig. 4.26 - Total Probability of Error: Gaussian-Gaussian Mixture
Variance 1, Signal Strength = 0.5

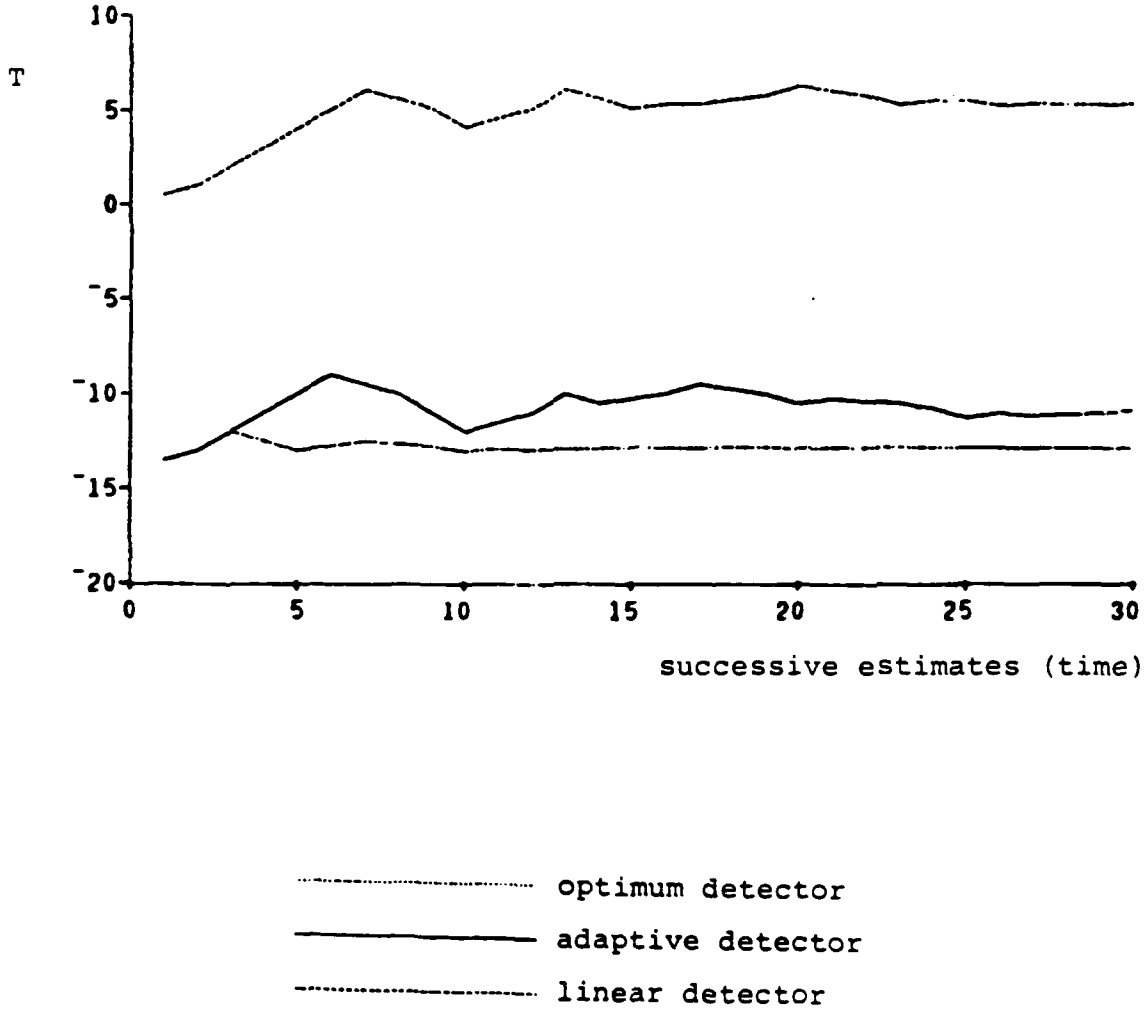


Fig. 4.27 - Threshold Levels: Gaussian-Gaussian Mixture
Variance 1, Signal Strength = 0.5

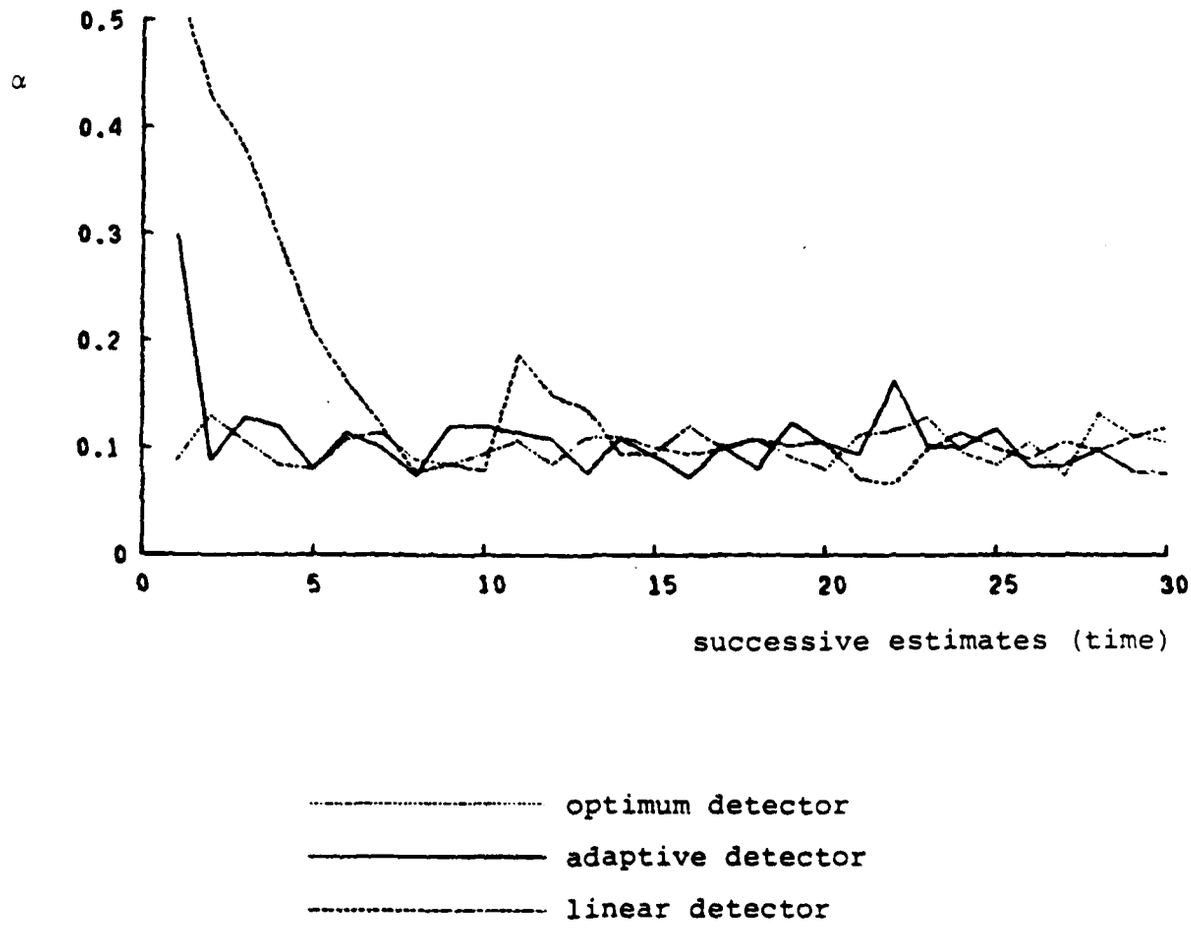


Fig. 4.28 - Probability of False Alarm: Gaussian-Gaussian Mixture
Variance 1, Signal Strength = 0.2

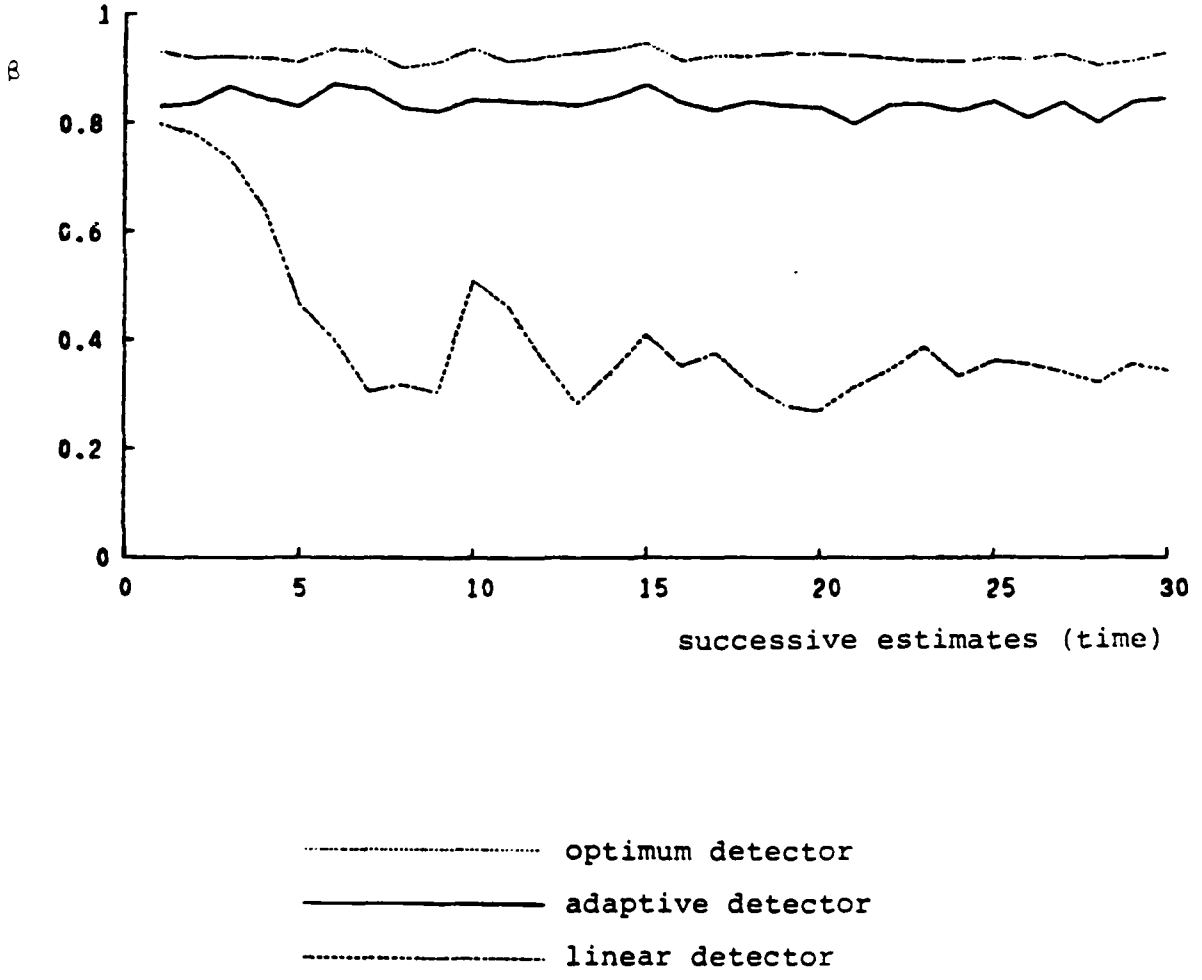


Fig. 4.29 - Probability of Detection: Gaussian-Gaussian Mixture
Variance 1, Signal Strength = 0.2

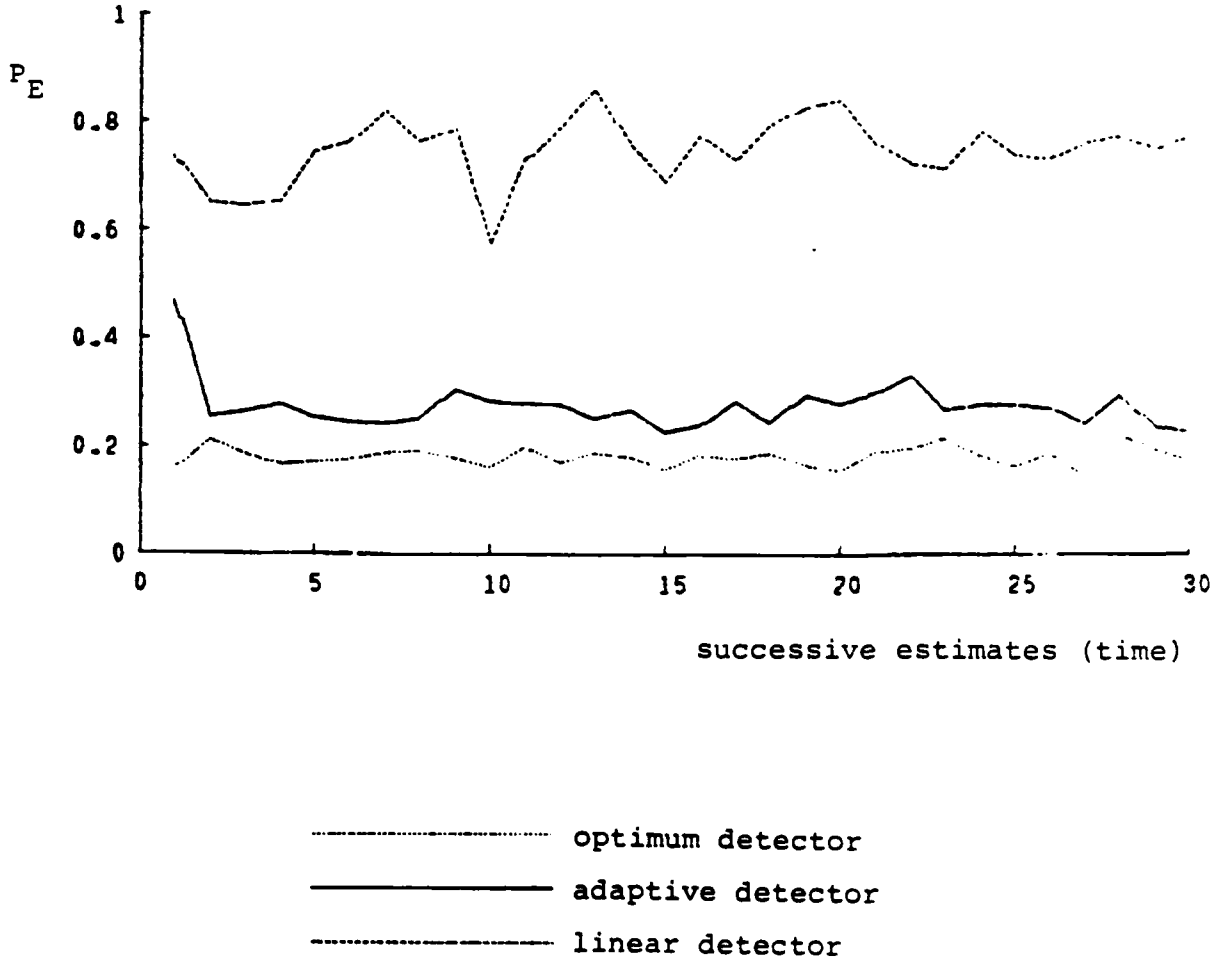


Fig. 4.30 - Total Probability of Error: Gaussian-Gaussian Mixture
Variance 1, Signal Strength = 0.2

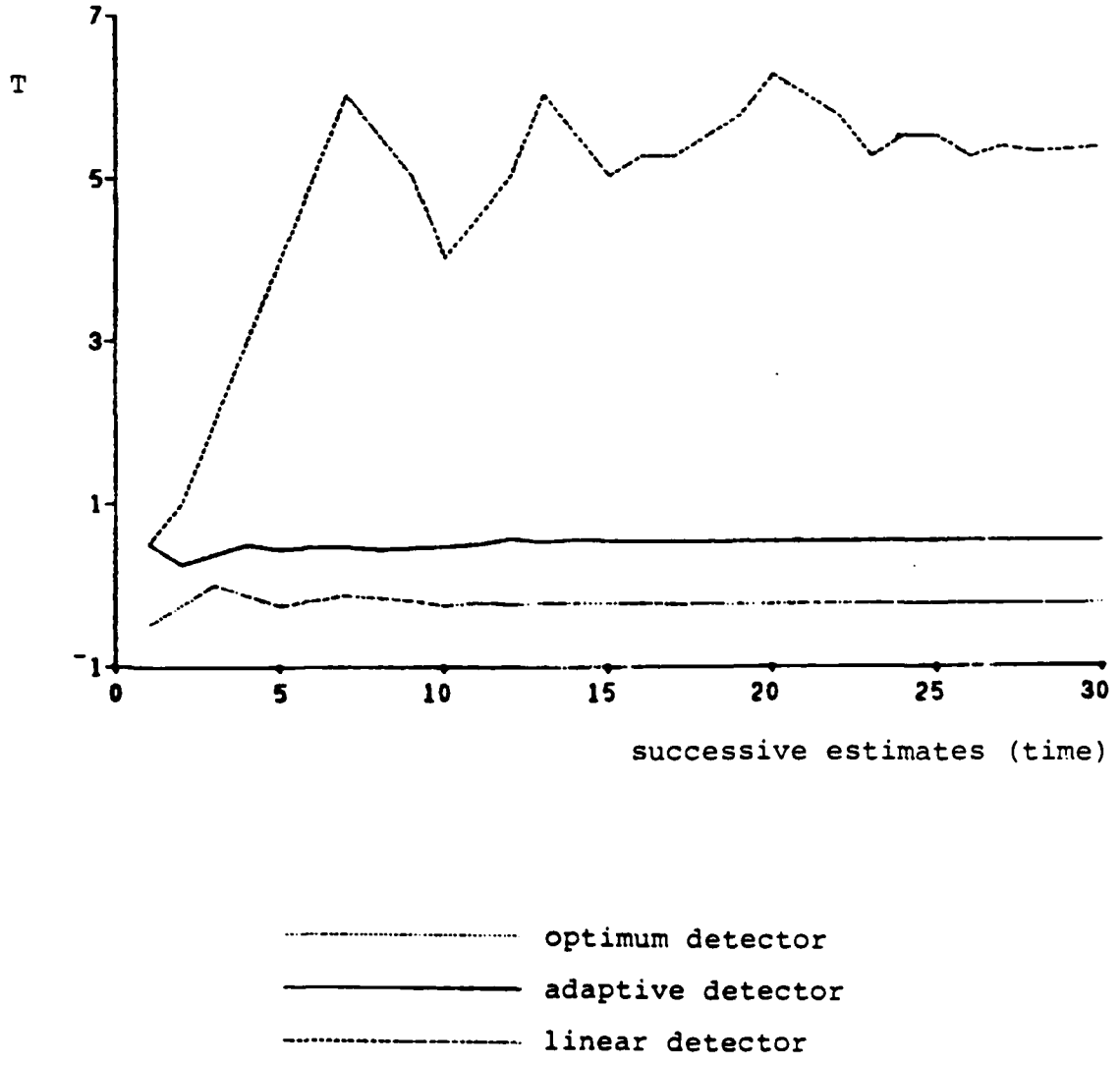


Fig. 4.31 - Threshold Levels: Gaussian-Gaussian Mixture
Variance 1, Signal Strength = 0.2

SIMULATION RESULTS

NOISE	DETECTOR	STEADY STATE MEAN α	STEADY STATE MEAN β	STEADY STATE MEAN TOTAL PROB. ERROR
GAUSSIAN	OPTIMAL	.097	.828	.269
	ADAPTIVE	.101	.832	.269
LAPLACE	OPTIMAL	.105	.945	.160
	ADAPTIVE	.108	.925	.183
	LINEAR	.098	.841	.257
JOHNSON (δ=1)	OPTIMAL	.095	.983	.112
	ADAPTIVE	.098	.982	.116
	LINEAR	.101	.862	.239
GAUSSIAN-GAUSSIAN γ^2=100 ϵ=0.1	OPTIMAL	.102	>.999	.102
	ADAPTIVE	.127	>.999	.127
	LINEAR	.108	.866	.242
GAUSSIAN-GAUSSIAN MIXTURE γ^2=100 ϵ=0.1 S=0.2	OPTIMAL	.102	.922	.180
	ADAPTIVE	.099	.833	.266
	LINEAR	.108	.347	.762

Fig. 4.32 - Simulation Results: α, β, P_E

However, both the optimum detector and the adaptive detector perform so well it is hard to differentiate between them. For this reason, this case was repeated with a decreased signal strength.

Using 0.2 for the signal rather than 0.5, the performances for these detectors are more readily apparent. Figs. 4.28 through 4.31 display α, β, P_E and T. Note that the scale on the vertical axis for both Figs. 4.29 and 4.30 have been changed. The linear detector's performance is now totally unacceptable. However, the adaptive detector's performance is close to that of the optimum detector, even with this relatively small signal.

Fig. 4.32 displays the average value of α, β and the total probability of error. In computing this average the first ten estimates were ignored to allow the threshold to stabilize. Due to the fact that the estimated value for α did deviate from 0.1, the estimates of β should not be compared blindly. The column for total probability of error removes this variability and is probably the best column for comparison purposes. In all cases the adaptive detector is found to be nearly optimum. For the heavier tailed densities it is also found to perform far better than the linear detector.

A Verification of Some Results from Chapter III

The results of Chapter III - Section 1 can be used to find asymptotic values (small signal, large numbers of observations) for the total probability of error for optimum detectors. The

table in Fig. 4.33 displays the five cases which were simulated. Adding the value of 0.1 for α to $1 - \beta$ from Eq. (3.13) yields the asymptotic probability of error for the optimum detector. The final column lists the probability of error observed from simulations. These numbers are in close agreement. This verifies both Eq. (3.13) and the accuracy of the simulation results. Note that the total probability of error for densities with high values for Fisher's Information is significantly less than for densities with low values for Fisher's Information. This supports the conclusion from Chapter III that Fisher's Information can serve as a measure of the difficulty of detecting an additive signal in noise. Figs. 4.34 through 4.42 also serve to illustrate this point.

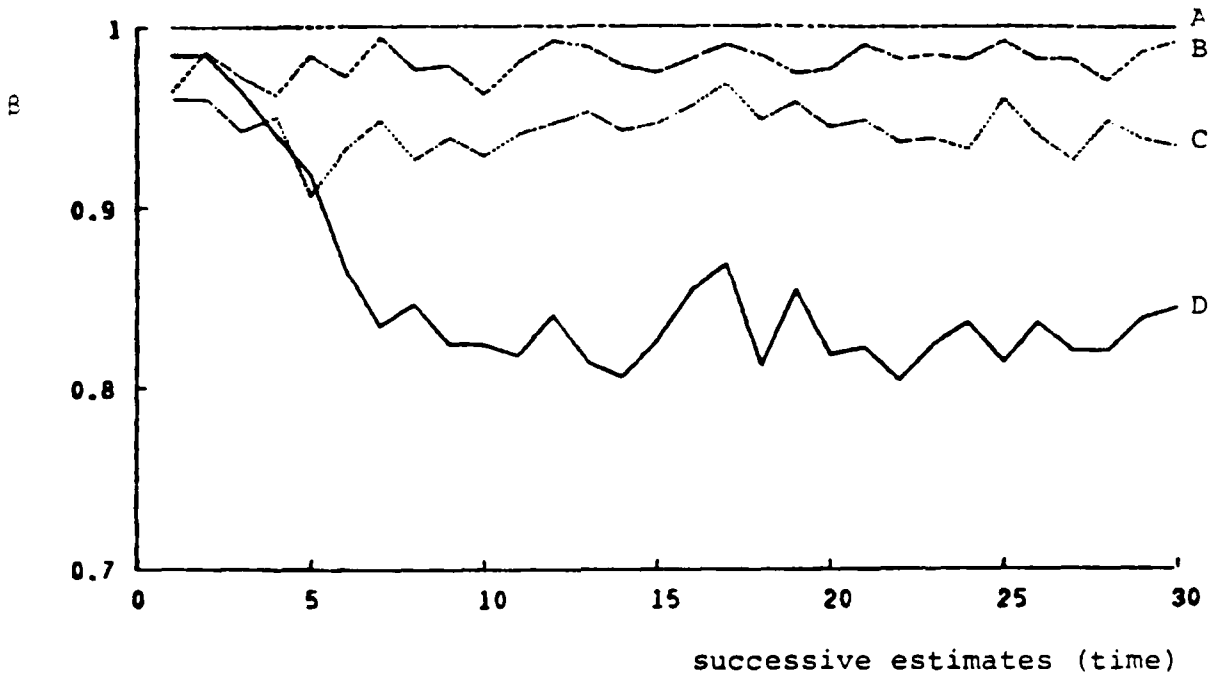
Fig. 4.34 displays the probability of detection (β) for optimum detectors in Gaussian, Laplace, Johnson ($\sigma = 1$), and Gaussian-Gaussian mixture noise. Fig. 4.35 displays β for the adaptive detector and Fig. 4.36 displays β for the linear detector. For both adaptive and optimum detectors, densities with high values for Fisher's Information yield higher power than densities with low values for Fisher's Information. The linear detector behaves about the same for all cases. One can think of the linear detector as being uniformly poor for all non-Gaussian densities.

Figs. 4.37, 4.38 and 4.39 display both α and β for the optimum, adaptive and linear detectors. These curves are included to show all of the data plotted to the same scale. Figs. 4.40, 4.41 and 4.42 display the probability of error for optimum, adaptive, and linear detectors. The same observations

Noise	Fisher's Information	Calculated Asymptotic* Probability of Error	Observed* Probability of Erro
Gaussian	1	.270	.269
Laplace	2	.130	.160
Johnson (δ=1)	2.386	.115	.112
Gaussian-Gaussian Mixture	9.02	.100	.102
Gaussian-Gaussian Mixture (signal=0.2)	9.02	.180	.180

* Values reported to 3 significant digits only

Fig. 4.33 - Comparison of Computed and Observed Probability of Error for Optimum Detectors



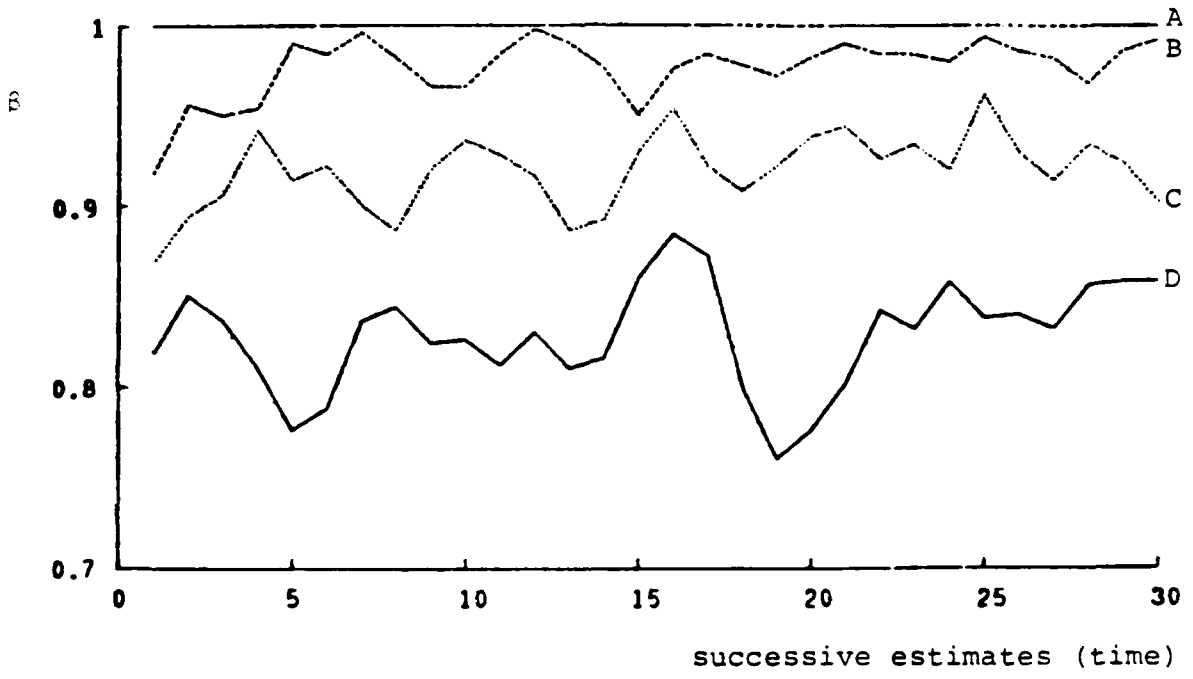
A ---------- Gaussian-Gaussian Mixture

B ---------- Johnson Noise, $\delta = 1$

C ---------- Laplace Noise

D ---------- Gaussian Noise

Fig. 4.34 - Probability of Detection: Optimum Detectors



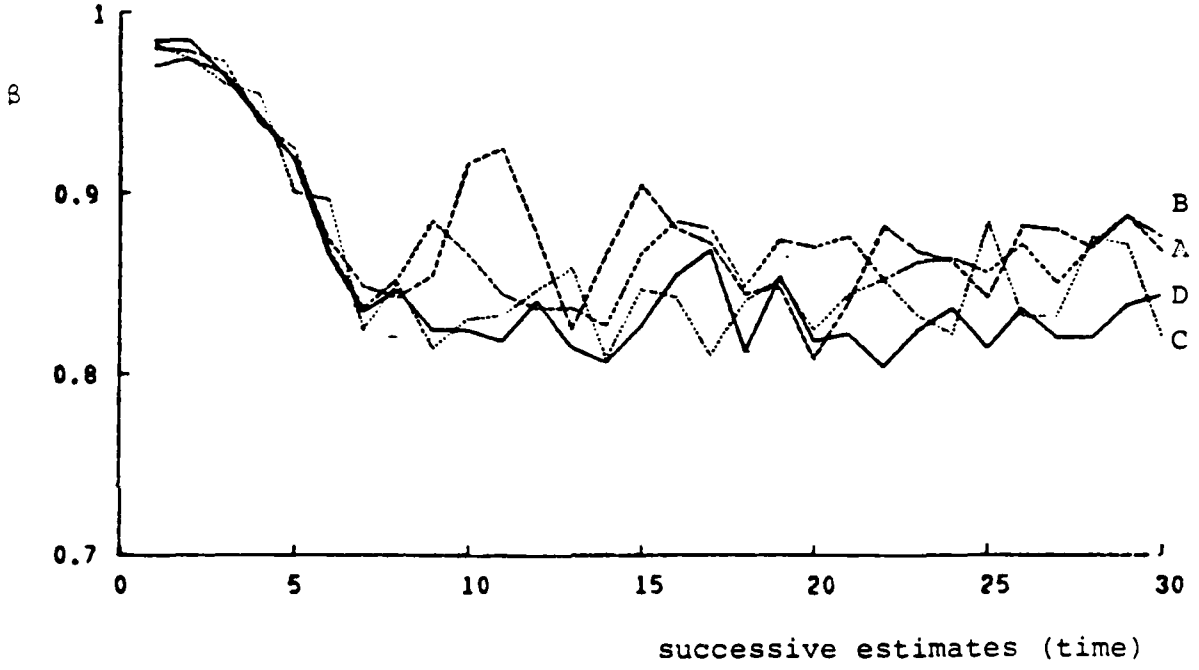
A ---------- Gaussian-Gaussian Mixture

B ---------- Johnson Noise, $\delta = 1$

C ---------- Laplace Noise

D ---------- Gaussian Noise

Fig. 4.35 - Probability of Detection: Adaptive Detector



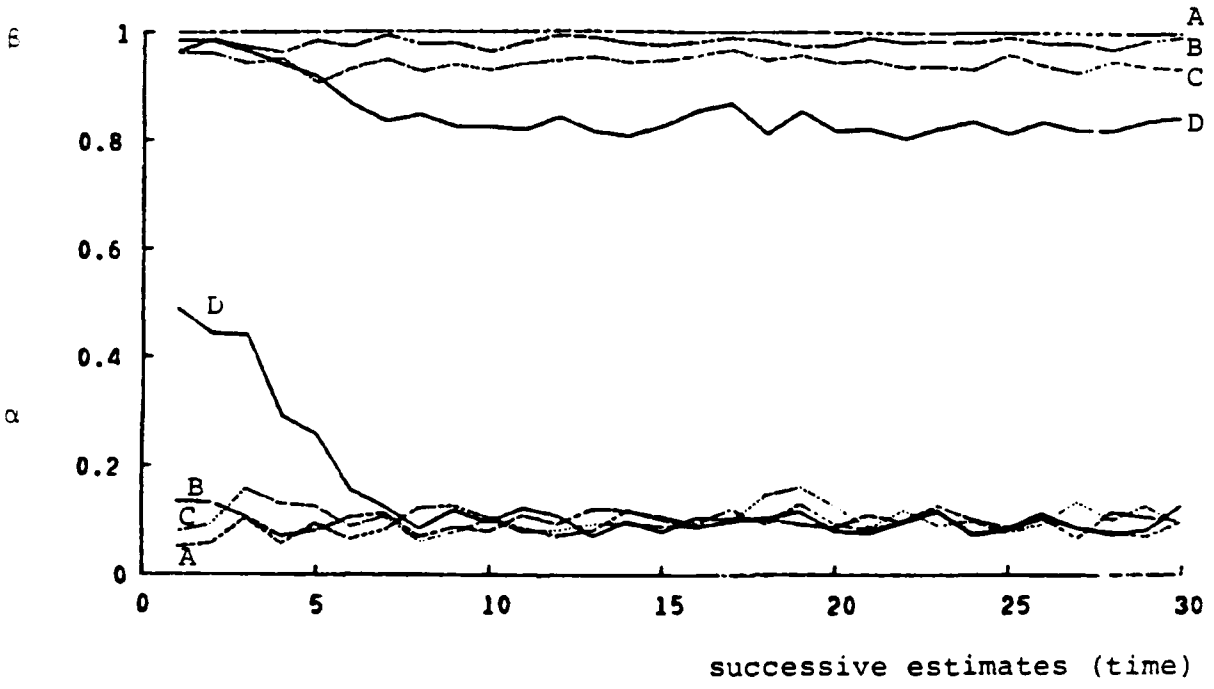
A ---------- Gaussian-Gaussian Mixture

B ---·---·--- Johnson Noise, $\delta = 1$

C ---·····--- Laplace Noise

D ——————— Gaussian Noise

Fig. 4.36 - Probability of Detection: Linear Detector



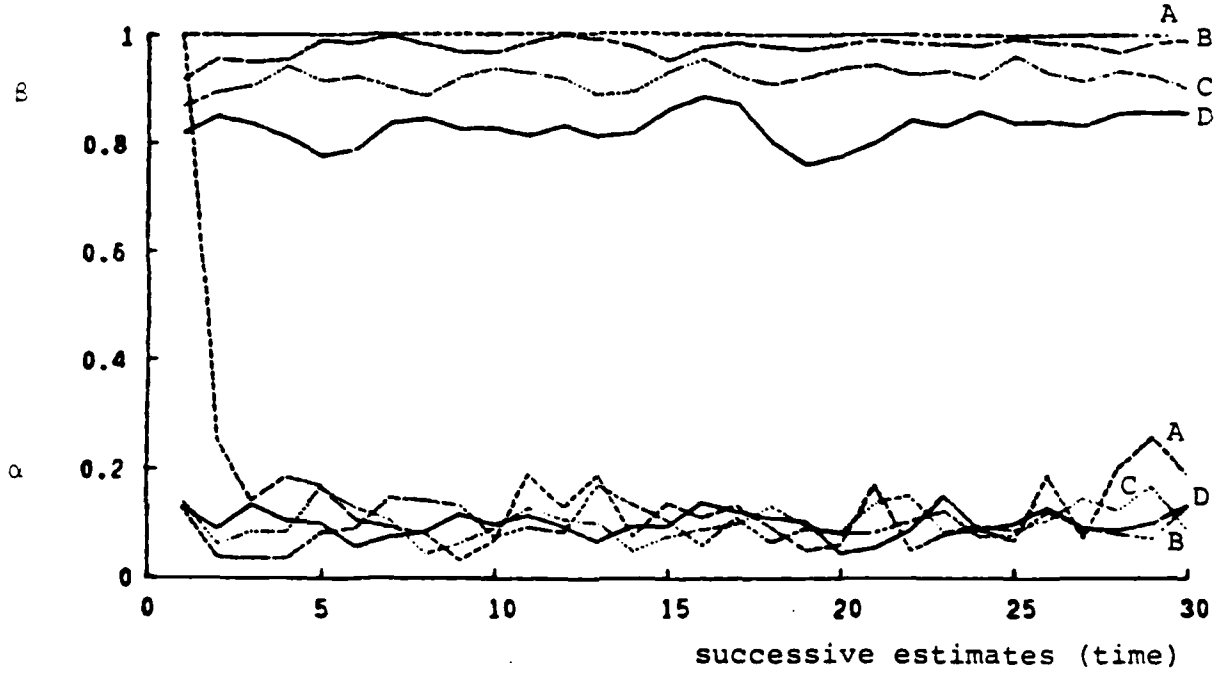
A ---------- Gaussian-Gaussian Mixture

B - - - - - - Johnson Noise, $\delta = 1$

C ——— Laplace Noise

D ——— Gaussian Noise

Fig. 4.37 - α and β: Optimum Detectors



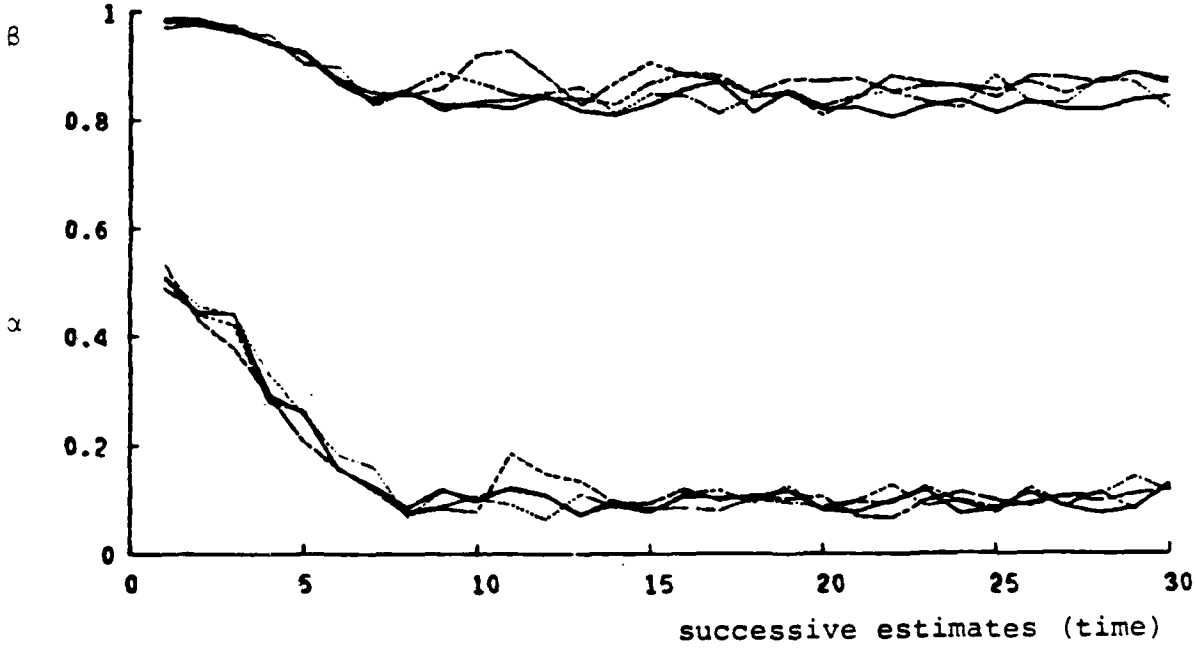
A ———————— Gaussian-Gaussian Mixture

B ———·———·——— Johnson Noise, $\delta = 1$

C ———·····——— Laplace Noise

D ———————— Gaussian Noise

Fig. 4.38 - α and β: Adaptive Detector



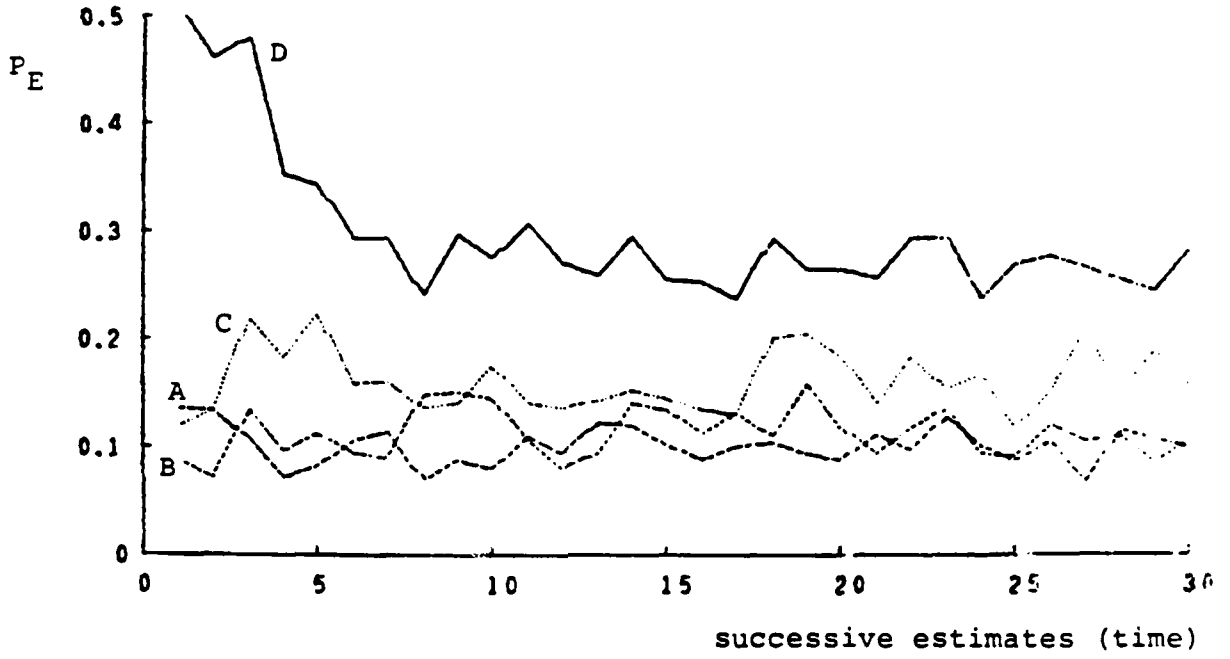
Gaussian-Gaussian Mixture

Johnson Noise, δ = 1

Laplace Noise

Gaussian Noise

Fig. 4.39 - α and β: Linear Detector



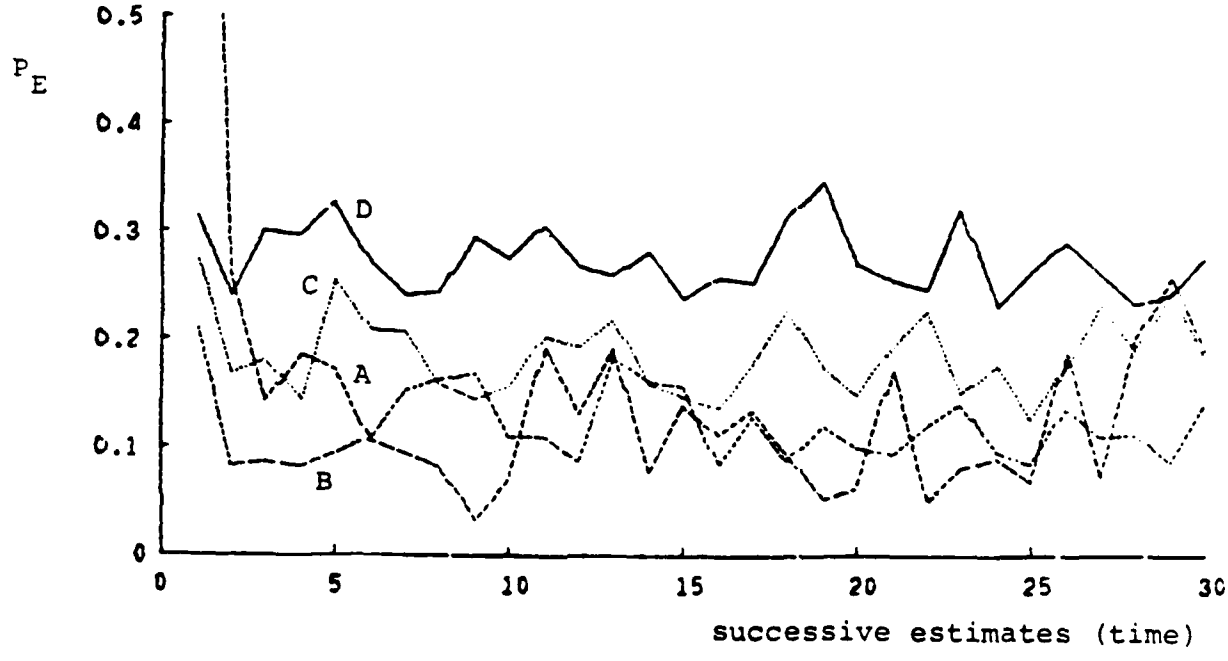
A ———————— Gaussian-Gaussian Mixture

B — — — — — — Johnson Noise, $\delta = 1$

C ——·——·—— Laplace Noise

D ———————— Gaussian Noise

Fig. 4.40 - Total Probability of Error: Optimum Detectors



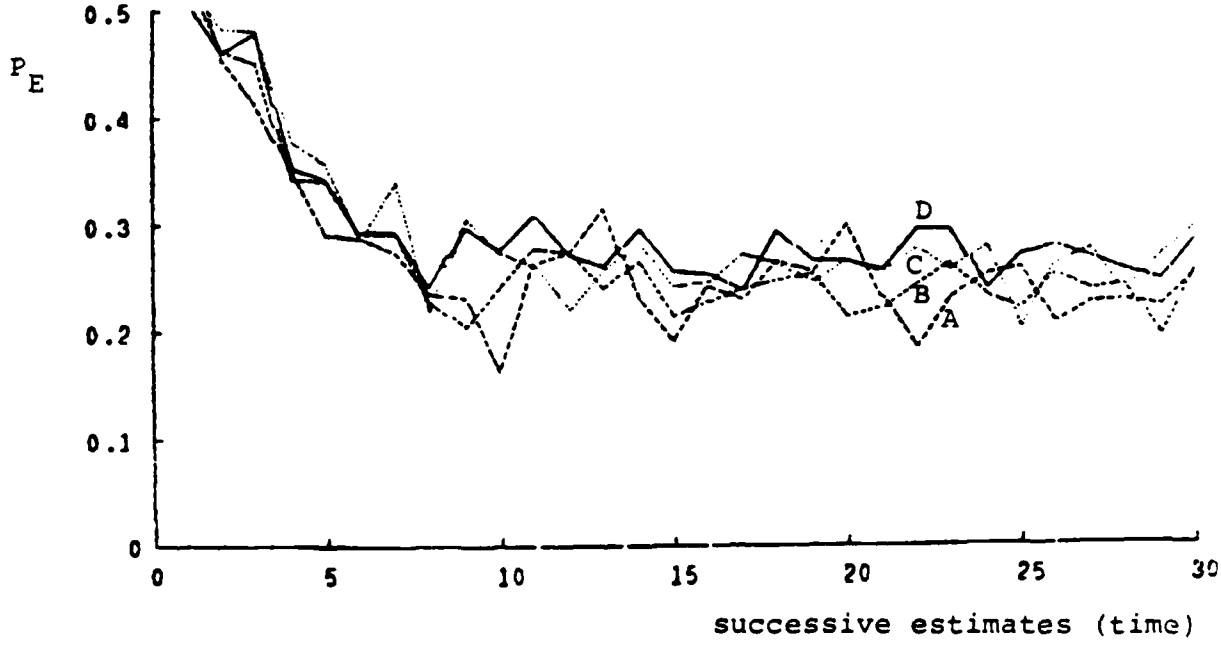
A ---------- Gaussian-Gaussian Mixture

B ---·---·--- Johnson Noise, $\delta = 1$

C ---·····--- Laplace Noise

D ——————— Gaussian Noise

Fig. 4.41 - Total Probability of Error: Adaptive Detector



A ---------- Gaussian-Gaussian Mixture

B ---------- Johnson Noise, $\delta = 1$

C ---------- Laplace Noise

D ---------- Gaussian Noise

Fig. 4.42 - Total Probability of Error: Linear Detector

which were made for the power curves can be made here. Highly non-Gaussian noise, with high values for Fisher's Information, can be an asset since the increased structure of the noise can be utilized to increase detector performance over that possible with just Gaussian noise. Nonlinear processing, however, is required to achieve this improvement. The linear detector's performance remained relatively the same for all densities studied.

The Performance of the Sign Detector

The sign detector was also considered in this simulation study. The sign detector or hard limiter may be expressed as

$$T(\underline{x}) = \sum_{i=1}^{N} \operatorname{sgn}(x_i) \tag{4.9}$$

where,

$$\operatorname{sgn}(x) = \begin{cases} +1 & , x>0 \\ 0 & , x=0 \\ -1 & , x<0 \end{cases}$$

Figs. 4.43 through 4.46 display α, β, P_E and the threshold for the sign detector in both Gaussian and Laplace noise. As before, N = 20 samples were used. In both cases, although the threshold had stabilized near 6, the curves for α and β cycled between two levels. To understand this, α and β were calculated numerically.

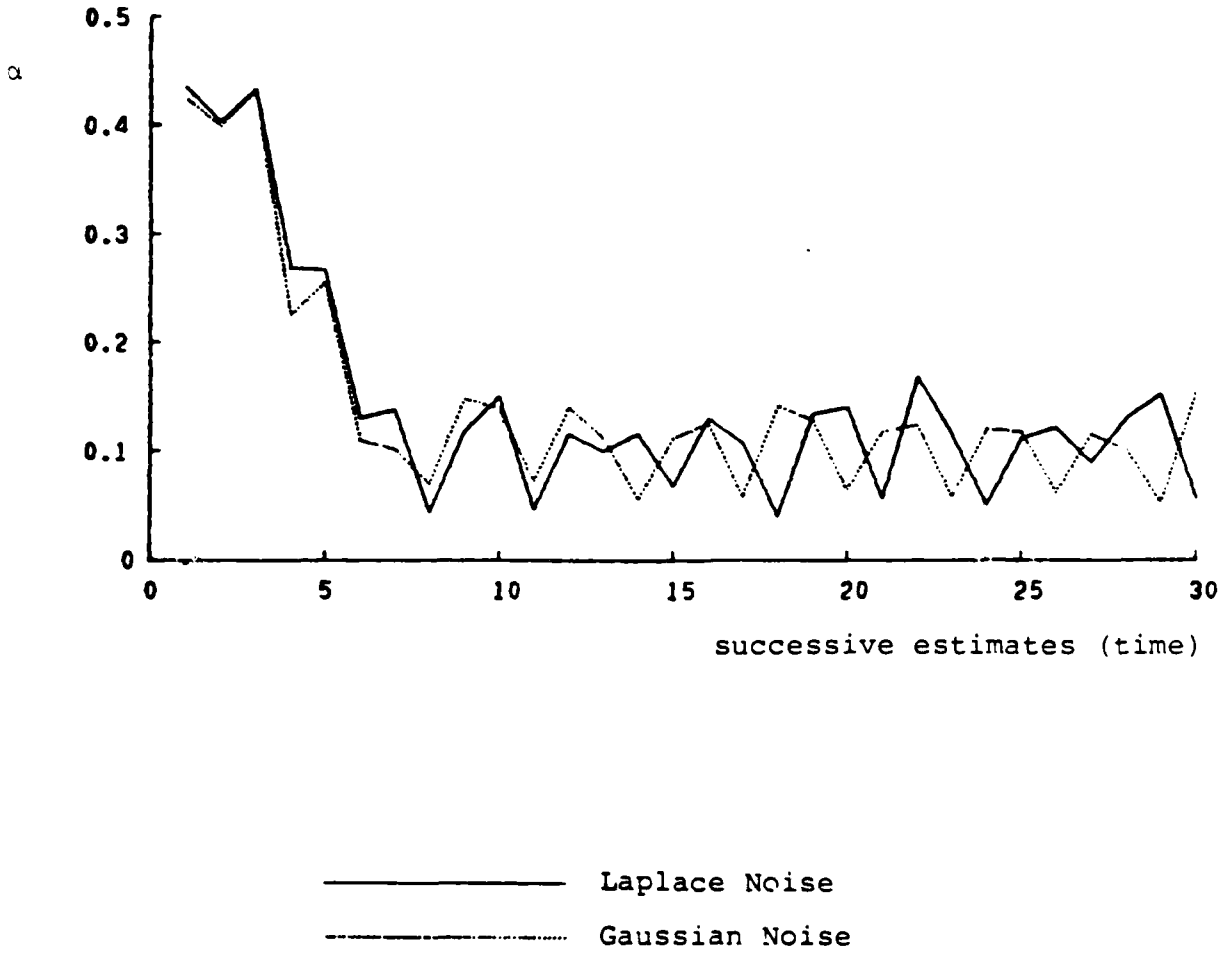


Fig. 4.43 - Probability of False Alarm: Sign Detector

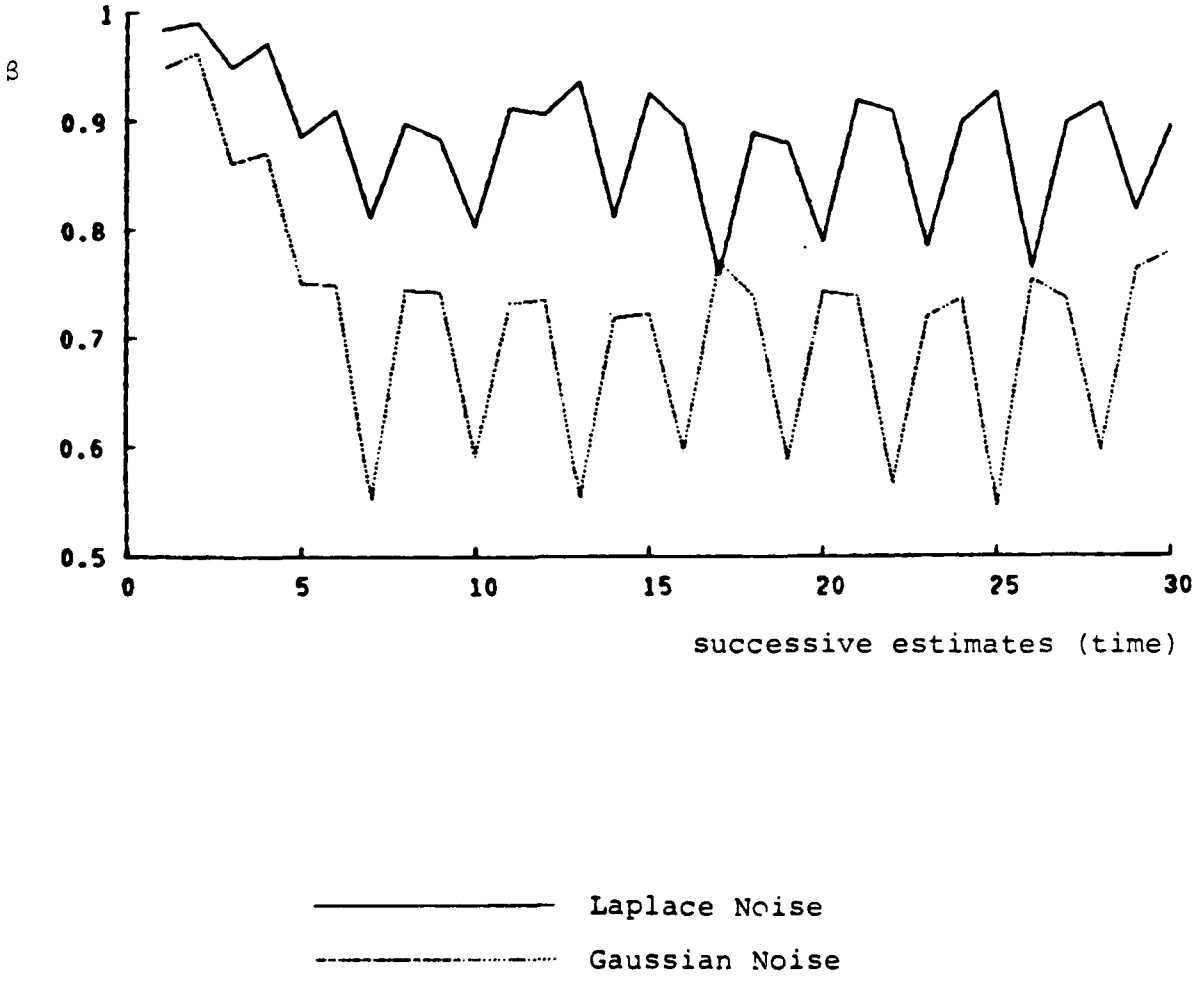


Fig. 4.44 - Probability of Detection: Sign Detector

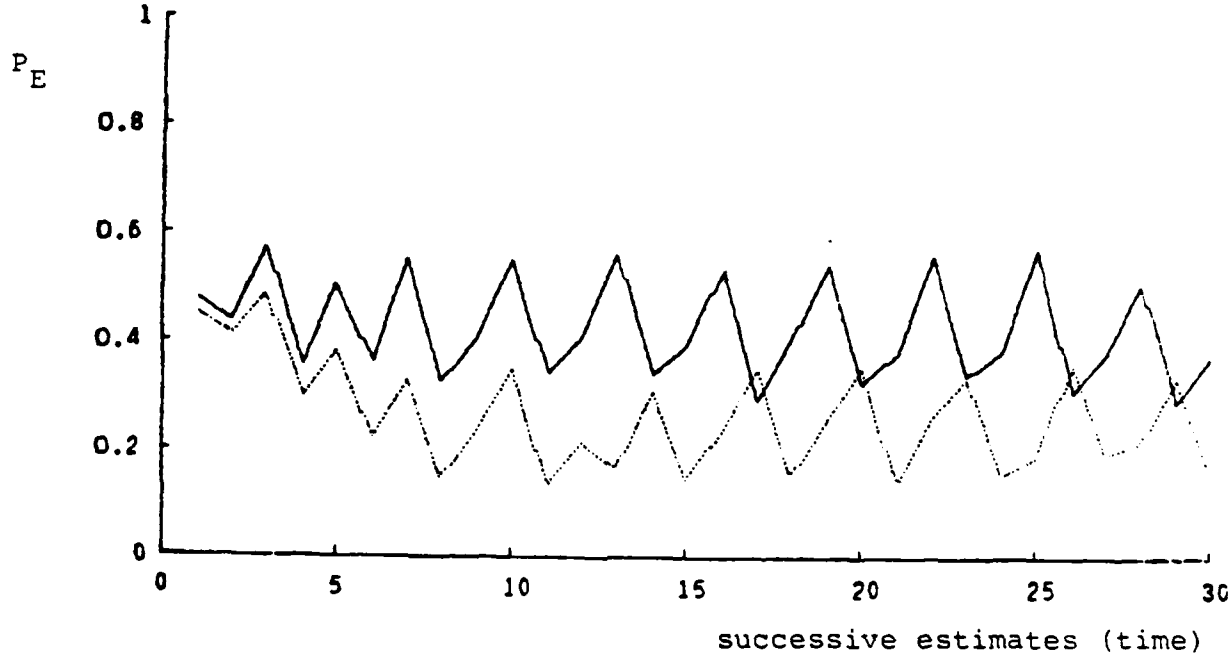
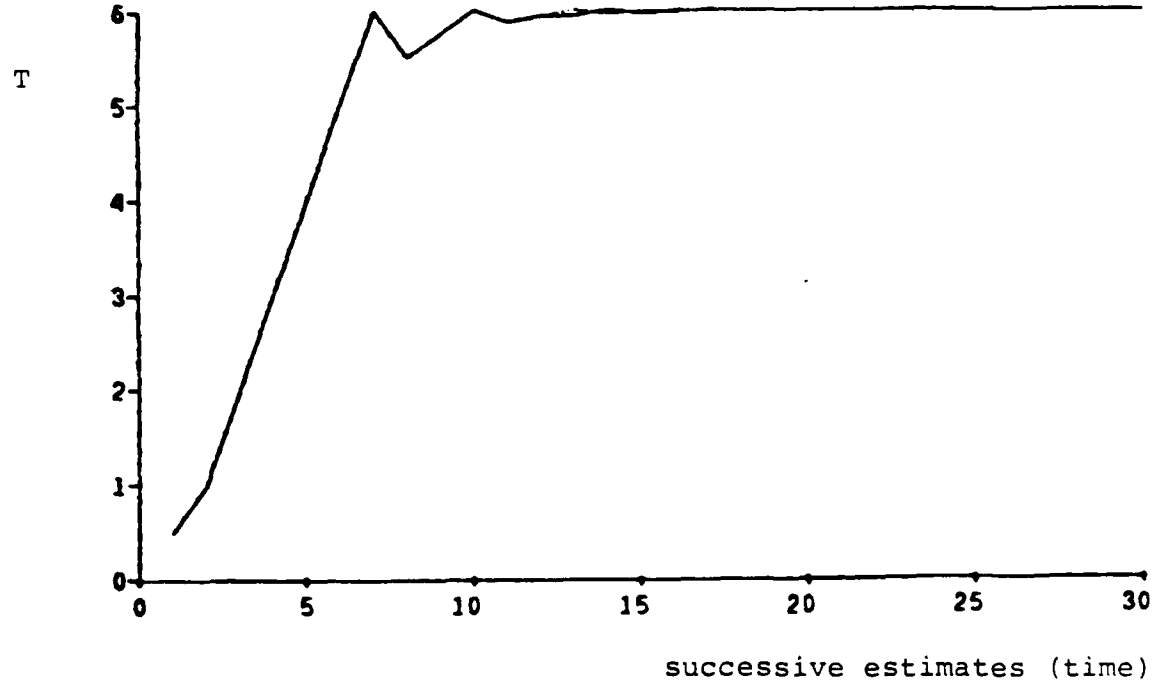


Fig. 4.45 - Total Probability of Error: Sign Detector



——————— Laplace Noise

– – – – – – Gaussian Noise

Fig. 4.46 - Threshold Levels: Sign Detector

Under H, when the noise has a symmetric distribution, $\text{sgn}(x_i)$ will be -1 with probability 1/2 and +1 with probability 1/2. Considering the sum of observations as binomial trials, the probability distribution for $T(\underline{x})$ can be calculated. It consists of impulses at all even numbers from -20 to 20 as shown in Fig. 4.47. Summing these impulses it can be seen that

$$\begin{aligned} \alpha = \text{Prob}_H[T(\underline{x}) > T] &= 0.132 \quad \text{for } 4 \leqslant T < 6 \\ &= 0.058 \quad \text{for } 6 \leqslant T < 8 \end{aligned} \tag{4.10}$$

The oscillations in α are due to the following reasons. Whenever $4 \leqslant T < 6$, the detector finds $\hat{\alpha} \approx 0.132$. The threshold is then raised slightly in an attempt to lower $\hat{\alpha}$ to 0.1. As soon as T reaches 6 ($6 \leqslant T < 8$), $\hat{\alpha}$ falls to approximately 0.058. Since $\alpha = 0.1$ is impossible to achieve, the system continues to oscillate with T very close to 6. A corresponding oscillation in β occurs.

As a check, β for Gaussian noise can be calculated easily. For a signal of 1/2, $\text{sgn}(x_i)$ equals -1 with probability 0.3085 and +1 with probability 0.6915. The distribution of $T(\underline{x})$ is given by Fig. 4.48. For $4 \leqslant T < 6$, $\beta = .746$. When $6 \leqslant T < 8$, $\beta = .575$. These numbers are in close agreement with those actually observed (Fig. 4.44).

These oscillations are a serious drawback for the sign detector. Probably the best way to eliminate this problem is to utilize a randomized test. Such tests have not been considered here since the sign detector is the only detector mentioned above which yields outputs which are discrete random variables.

t	$\text{Prob}_H[T(\underline{x}) = t]$
20	9.54×10^{-7}
18	1.91×10^{-5}
16	1.81×10^{-4}
14	1.09×10^{-3}
12	4.62×10^{-3}
10	1.48×10^{-2}
8	3.70×10^{-2}
6	7.39×10^{-2}
4	1.20×10^{-1}
2	1.60×10^{-1}
0	1.76×10^{-1}

For -2 to -20, $\text{Prob}[T(\underline{x}) = -t] = \text{Prob}[T(\underline{x}) = t]$

Fig. 4.47 - Probability Distribution of the Output of the Sign Detector: Symmetric Noise Density, No Signal

t	$\text{Prob}_K[T(\underline{x}) = t]$
20	6.249×10^{-4}
18	5.576×10^{-3}
16	2.363×10^{-2}
14	6.326×10^{-2}
12	1.199×10^{-1}
10	1.712×10^{-1}
8	1.910×10^{-1}
6	1.704×10^{-1}
4	1.235×10^{-1}
2	7.349×10^{-2}
0	3.606×10^{-2}
-2	1.463×10^{-2}
-4	4.894×10^{-3}
-6	1.344×10^{-3}
-8	2.997×10^{-4}
-10	5.349×10^{-5}
-12	7.457×10^{-6}
-14	7.827×10^{-7}
-16	5.820×10^{-8}
-18	2.733×10^{-9}
-20	6.097×10^{-11}

Fig. 4.48 - Probability Distribution of the Output of the Sign Detector: Gaussian Noise, Signal = 1/2

References - Chapter IV

1. J. J. Filliben, "Simple and Robust Linear Estimation of the Location Parameter of a Symmetric Distribution," Ph.D. dissertation, Dept. of Statistics, Princeton Univ., Princeton, NJ, 1969.

Chapter V

CONCLUSION

In this dissertation, methods for the detection of known, discrete-time signals in impulsive noise have been investigated. After a brief introduction to the detection problem in Chapter I, Chapter II concentrated on impulsive noise. The characteristics and sources of impulsive noise were considered briefly. Several impulsive noise models were explained. Three systems of density functions were then presented to provide first order non-Gaussian noise models which are simple enough to use in detection problems. These three systems are: a generalized Gaussian noise, the Johnson S_u System, and a mixture model. Further research is needed in developing impulsive noise models which are tractable in detection problems. Middleton's model [1] is probably the most complete and comprehensive to date. However, it is quite complex and is unable to model the dependency which has been observed for impulsive noise samples. Research concentrating on modeling this dependency structure and on developing higher order densities for non-Gaussian noise is needed.

In Chapter III, the importance of Fisher's Information in assessing the asymptotic performance of detectors has been stressed. It has been shown that an impulsive noise with a

high value for Fisher's Information provides a better environment for detection than does a Gaussian noise of equal power. The increased structure of impulsive noise may be exploited to increase detector performance over that possible for Gaussian noise. Nonlinear processing, however, is required to achieve this improvement. To further illuminate the relationship between Fisher's Information and detection, it may be beneficial to explore the connection between Fisher's Information and entropy. To illustrate this connection, observe that among all densities of a given variance, the Gaussian distribution both minimizes Fisher's Information and maximizes entropy. Further research centered on exploring this connection and on the effect of these measures on the analysis of detector performance would be desirable.

The relationship between Fisher's Information and asymptotic detector performance was utilized in Chapter III for the design of minimax detectors. The density with minimum Fisher's Information over a given class of densities is the worst case density within the class for detection. The optimum detector for this density is the minimax detector for the class. Several density classes were defined and conditions for the minimax detectors were explored. One especially interesting class is the class of all distributions whose cumulative distribution functions pass through k given points. Huber's conditions [2] which define the density that minimizes Fisher's Information over this class were presented. This technique appears especially promising for detection purposes. A possible implementation would require taking data points and

finding the density with minimum Fisher's Information which passes through the points. Then, the optimum detector for this density can be used. This detection procedure should be quite robust.

Minimax techniques can be useful if the class of densities is properly defined. However, in many cases it may be more appropriate to find a detector which works well for many non-Gaussian densities rather than to find the one that works best for the worst case density. To approach this problem one could examine the equation for efficacy [Eq. (3.39)] for various detectors and density systems, the objective being to find a nonlinearity which yields high values for efficacy for many different densities. This nonlinearity should then be tested by simulation to determine how well the performance predictions based on efficacy carry over to the small sample case.

In Chapter IV an adaptive detector based on the Johnson System was presented. Simulation results were displayed which showed that this detector performs well for several different densities. However, to truly test this system, actual impulsive noise data should be used. A variation of this detector could also be considered. The tail measure ρ, rather than τ, could be used to estimate the parameter δ. Use of ρ would require taking only one quantile. Since this point is on the tail of the density, fewer samples would have to be saved to find an estimate for the quantile. That is, when looking for the .995 quantile only the top 50 out of

10,000 points need be saved; while looking for the .975 quantile (required for estimating τ) requires saving 250 points. Fast methods to estimate both the quantiles and the rate of change of the quantiles [$dR(p)/dp$ where $R(p) = F^{-1}(p)$] are needed.

As briefly mentioned in Chapter IV, the tail measures β_2, τ, and ρ can be used as tools for the comparison of density systems. Different densities, whose tail behavior are identical as measured by either β_2, τ or ρ, could be compared in terms of Fisher's Information or the performance of various detectors. This study could aid in the selection of appropriate density systems for noise modeling.

References - Chapter V

1. D. Middleton, "Man-Made Noise in Urban Environments and Transportation Systems: Models and Measurements," IEEE Trans. Comm., Vol. COM-21, No. 11, pp. 1232-1241, November 1973.

2. P. J. Huber, "Fisher Information and Spline Interpolation," The Annals of Statistics, Vol. 2, No. 5, pp. 1029-1033, September 1974.

OFFICE OF NAVAL RESEARCH
STATISTICS AND PROBABILITY PROGRAM

BASIC DISTRIBUTION LIST
FOR
UNCLASSIFIED TECHNICAL REPORTS

FEBRUARY 1982

	Copies
Statistics and Probability Program (Code 411(SP)) Office of Naval Research Arlington, VA 22217	3
Defense Technical Information Center Cameron Station Alexandria, VA 22314	12
Commanding Officer Office of Naval Research Eastern/Central Regional Office Attn: Director for Science Barnes Building 495 Summer Street Boston, MA 02210	1
Commanding Officer Office of Naval Research Western Regional Office Attn: Dr. Richard Lau 1030 East Green Street Pasadena, CA 91101	1
U. S. ONR Liaison Office - Far East Attn: Scientific Director APO San Francisco 96503	1
Applied Mathematics Laboratory David Taylor Naval Ship Research and Development Center Attn: Mr. G. H. Gleissner Bethesda, Maryland 20084	1
Commandant of the Marine Coprs (Code AX) Attn: Dr. A. L. Slafkosky Scientific Advisor Washington, DC 20380	1
Navy Library National Space Technology Laboratory Attn: Navy Librarian Bay St. Louis, MS 39522	1
U. S. Army Research Office P.O. Box 12211 Attn: Dr. J. Chandra Research Triangle Park, NC 27706	1
Director National Security Agency Attn: R51, Dr. Maar Fort Meade, MD 20755	1
ATAA-SL, Library U.S. Army TRADOC Systems Analysis Activity Department of the Army White Sands Missile Range, NM 88002	1
ARI Field Unit-USAREUR Attn: Library c/o ODCSPER HQ USAEREUR & 7th Army APO New York 09403	1
Library, Code 1424 Naval Postgraduate School Monterey, CA 93940	1
Technical Information Division Naval Research Laboratory Washington, DC 20375	1
OASD (I&L), Pentagon Attn: Mr. Charles S. Smith Washington, DC 20301	1

Copies

Director
AMSAA
Attn: DRXSY-MP, H. Cohen
Aberdeen Proving Ground, MD 21005 1

Dr. Gerhard Heiche
Naval Air Systems Command (NAIR 03)
Jefferson Plaza No. 1
Arlington, VA 20360 1

Dr. Barbara Bailar
Associate Director, Statistical Standards
Bureau of Census
Washington, DC 20233 1

Leon Slavin
Naval Sea Systems Command (NSEA 05H)
Crystal Mall #4, Rm. 129
Washington, DC 20036 1

B. E. Clark
RR #2, Box 647-B
Graham, NC 27253 1

Naval Underwater Systems Center
Attn: Dr. Derrill J. Bordelon
Code 601
Newport, Rhode Island 02840 1

Naval Coastal Systems Center
Code 741
Attn: Mr. C. M. Bennett
Panama City, FL 32401 1

Naval Electronic Systems Command (NELEX 612)
Attn: John Schuster
National Center No. 1
Arlington, VA 20360 1

Defense Logistics Studies Information Exchange
Army Logistics Management Center
Attn: Mr. J. Dowling
Fort Lee, VA 23801 1

Copies

Reliability Analysis Center (RAC)
RADC/RBRAC
Attn: I. L. Krulac
Data Coordinator/
Government Programs
Griffiss AFB, New York 13441 1

Technical Library
Naval Ordnance Station
Indian Head, MD 20640 1

Library
Naval Ocean Systems Center
San Diego, CA 92152 1

Technical Library
Bureau of Naval Personnel
Department of the Navy
Washington, DC 20370 1

Mr. Dan Leonard
Code 8105
Naval Ocean Systems Center
San Diego, CA 92152 1

Dr. Alan F. Petty
Code 7930
Naval Research Laboratory
Washington, DC 20375 1

Dr. M. J. Fischer
Defense Communications Agency
Defense Communications Engineering Center
1860 Wiehle Avenue
Reston, VA 22090 1

Mr. Jim Gates
Code 9211
Fleet Material Support Office
U. S. Navy Supply Center
Mechanicsburg, PA 17055 1

Mr. Ted Tupper
Code M-311C
Military Sealift Command
Department of the Navy
Washington, DC 20390 1

	Copies		Copies
Mr. F. R. Del Priori Code 224 Operational Test and Evaluation Force (OPTEVFOR) Norfolk, VA 23511	1		